Contribution à la modélisation et à la gestion dynamique du risque des marchés de l'énergie

Noufel Frikha

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité :
Mathématiques Appliquées

présentée
par

Noufel Frikha

pour obtenir le grade de

DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de thèse:

Contribution à la modélisation et à la gestion dynamique du risque des marchés de l'énergie

Directeur de thèse: Gilles Pagès

Date de soutenance: 1er Décembre 2010

Rapporteurs

M. Bernard Lapeyre, Ecole des Ponts ParisTech
M. Fred Espen Benth, Université d'Oslo

Composition du Jury

M.	Bernard Lapeyre,	Rapporteur
Mme.	Nicole El Karoui,	Examinateur
M.	Jean Claude Fort,	Examinateur
M.	Damien Lamberton,	Examinateur
M.	Olivier Bardou,	Examinateur
M.	Fred Espen Benth,	Examinateur
M.	Gilles Pagès,	Directeur de Thèse
M.	Vincent Lemaire,	Co-directeur de Thèse

Remerciements

Bien loin d'être un pensum, ces remerciements sont avant tout pour moi l'occasion d'exprimer toute ma joie ainsi que toute ma sympathie aux personnes qui m'ont entouré.

Le soutien, l'humour, l'enthousiasme ainsi que l'encadrement scientifique très rigoureux de M. Gilles Pagès au cours de ces trois années de thèse effectuées sous sa direction m'ont permis d'aboutir à ce travail.

Par la rédaction de ces quelques lignes, le moment est donc venu pour moi de le remercier chaleureusement et de lui témoigner toute mon estime et ma sincère reconnaissance.

Je tiens à exprimer toute ma reconnaissance à Vincent Lemaire et Olivier Bardou avec qui j'ai beaucoup apprécié de collaborer et qui ont beaucoup contribué à l'accomplissement de cette thèse. Ce fut un véritable plaisir de travailler avec vous.

Je remercie Bernard Lapeyre et Fred Espen Benth qui ont accepté d'être les rapporteurs de cette thèse.

Mes remerciements vont aussi aux autres membres du Jury: Damien Lamberton, Nicole El Karoui, Jean-Claude Fort.

Je suis très honoré qu'ils aient accepté d'y faire partie. Plus particulièrement, je remercie vivement Mme Nicole El Karoui pour la qualité de son enseignement au master Probabilités et Finance, Damien Lamberton et Bernard Lapeyre dont la clarté et la rigueur de leur livre *Introduction au calcul stochastique appliqué à la finance* m'ont beaucoup servi durant ces trois années ainsi que Jean-Claude Fort pour ses remarques qui m'ont permis d'améliorer la qualité de ce manuscrit.

J'exprime également toute ma considération à tous les membres du Laboratoire de Probabilités et de Modèles Aléatoires au sein duquel j'ai effectué ce travail, plus particulièrement les personnes avec qui j'ai eu l'occasion de discuter au gré des déjeuners ou plus simplement lors de mes rencontres dans les couloirs. Je pense notamment à M. Portes dont la disponibilité en cas de problèmes concernant l'informatique et la bonne humeur m'ont beaucoup touché. Mes pensées vont aussi aux personnes avec qui j'ai partagé le même bureau: Mathieur Richard, Raoul Normand, Eric Lucon, Sophie Dede, Stavros Vakeroudis, Nikos, Joachim Lebovits, Sylvain Corlay et Camille Illande, sans oublier Sophie Laruelle et ceux du bureau 16-26.225. Je vous remercie pour tous ces bons moments.

Je voudrais maintenant remercier chaleureusement les membres du Projet Sirius-Haltea et du Projet Prix de la Direction de la Recherche et de l'Innovation de GDFSuez pour toutes les discussions au sujet des marchés de l'énergie, votre collaboration et votre sympathie. Notamment, Olivier Bardou qui m'y a accueilli, Sandrine Bouthemy et Céline Jerusalem qui par leurs remarques m'ont beaucoup apporté. Mais aussi les autres membres des équipes et je ne me risquerai pas à tous

vous citer de peur d'en oublier un. J'ai beaucoup appris grâce à vous et je vous en remercie.

A toute ma famille dont l'amour indéfectible m'a porté et soutenu jour après jour et sans laquelle je ne serai probablement pas arrivé jusqu'içi aujourd'hui, c'est à mon tour de vous exprimer ma profonde reconnaissance. Tout d'abord à ma mère et à mon défunt père, je vous remercie pour tout. Puis à ma soeur et son mari, ainsi qu'à mon frère, vous avez beaucoup fait pour moi.

Je termine en remerciant mes quelques amis, plus particulièrement ceux du cours de théâtre Jean-Laurent Cochet. Lorsque je m'évoque tous ces bons moments, tous ces bons souvenirs, frangés de nostalgie, je me dis que ce travail vous doit également. Enfin à ma douce et tendre Elise, je tiens à te remercier très chaleureusement pour ton amour qui m'a tendrement accompagné au cours de cette thèse.

Contents

Chapter 1

Introduction

Cette thèse porte sur l'application des algorithmes stochastiques à la gestion et au contrôle des risques sur les marchés financiers et de l'énergie. Rares sont les activités économiques (ou dans d'autres domaines comme la physique, la biologie, ...) qui ne comportent aucun risque. Dans le cadre des activités des marchés financiers et de l'énergie on en distingue principalement quatre:

- Le risque de crédit (ou de contrepartie) résulte de l'incertitude qu'un tiers (un particulier ou une entreprise) ne rembourse pas sa dette à l'écheance fixée.
- Le risque de marché résulte de l'exposition de la valeur d'un portefeuille aux fluctuations des facteurs de risque liés au marché (cours des actions, taux, prix des matières premières, ...)
- Le risque climatique provient de la dépendance de la valeur d'un portefeuille aux aléas climatiques (température, intensité du vent, ...)
- Le risque opérationnel est le risque qu'une erreur humaine, une panne (turbine à gaz, stockage gazier,...), un dysfonctionnement (informatique) perturbe l'activité économique du détenteur du portefeuille.

Ce travail se compose de trois chapitres indépendants.

Le premier chapitre développe une méthode d'estimation par algorithme stochastique de deux mesures de risque couramment utilisées dans la pratique du contrôle des risques: la *Value-at-Risk* (VaR) et la *Conditional Value-at-Risk* (CVaR). Étant fixé un niveau de confiance $\alpha \in (0,1)$ et un horizon de temps T (la valeur $\alpha = 95\%$ est souvent utilisée), la VaR au seuil α (VaR_α) de la perte du portefeuille est la valeur qui ne sera pas dépassée par celle-ci avec une probabilité α. Avec une telle définition, la VaR est un indicateur synthétique du risque utile dans beaucoup d'autres domaines que la finance: physique, biologie, assurance,...

Pour autant, la VaR ne donne aucune information sur la perte maximale possible du portefeuille. Les mouvements des marchés sont parfois tels que les pertes peuvent dépasser le seuil de la VaR et s'avérer très lourdes si la queue de distribution des pertes du portefeuille est épaisse. La CVaR au seuil α tente de palier ce manque. Elle représente la moyenne des pertes du portefeuille lorsque celles-ci sont supérieures à la VaR_α. Ainsi, la VaR et la CVaR sont fortement liées. Elles le sont

d'autant plus qu'elles apparaissent comme la solution et la valeur d'un même problème d'optimisation convexe (c.f. [78] et [77]). La fonction objective de ce problème et son gradient s'écrivent sous forme d'espérance. Par conséquent, l'estimation de ces deux quantités peut être réalisée à l'aide d'un algorithme stochastique.

Un algorithme stochastique est une suite de vecteurs aléatoires $(X_n)_{n\geq 1}$ prenant ses valeurs dans un espace euclidien de dimension finie (en général dans $\mathbb{R}^d$) et définie récursivement par

$$\forall n \in \mathbb{N}, \quad X_{n+1} = X_n + \gamma_{n+1} V_{n+1}, \tag{1.1}$$

où la suite positive $(\gamma_n)_{n\geq 1}$, appelé pas de l'algorithme vérifie $\sum_{n\geq 1} \gamma_n = +\infty$ et où V_{n+1} est une variable aléatoire appelée l'observation de l'algorithme dépendant de la variable X_n, qui peut par exemple caractériser l'état d'un système au pas n dont son évolution au pas $n+1$ dépend de son propre état et d'une variable aléatoire ϵ_{n+1} de loi μ (indépendante de X_0), *i.e.* $V_{n+1} = H(X_n, \epsilon_{n+1})$, avec $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ vérifiant:

$$\forall x \in \mathbb{R}^d,\ H(x,) \in L^1_{\mathbb{R}^d}(\mu) \quad \text{et} \quad h : x \mapsto \int_{\mathbb{R}^q} H(x,u)\mu(\mathrm{d}u) \ \text{ est borélienne.}$$

A partir d'une observation ou d'une simulation, l'utilisateur disposant de X_n construit $H(X_n, \epsilon_{n+1})$, *i.e.* la variable aléatoire ϵ_{n+1} de loi μ est simulable à un coût raisonnable et la fonction H est calculable numériquement à un coût raisonnable également. La fonction moyenne de l'algorithme h n'est pas calculable numériquement à un coût raisonnable. Si elle l'est, on remplacera l'algorithme stochastique ci-dessus par son homologue déterministe

$$x_{n+1} = x_n - \gamma_{n+1} h(x_n), \quad n \geq 0.$$

Par conséquent, il est courant d'utiliser un algorithme stochastique lorsque le calcul numérique de h est couteux comparé au calcul numérique de H et de la simulation de la loi μ. Remarquons que l'algorithme stochastique (1.1) peut s'écrire

$$\forall n \in \mathbb{N}, \quad X_{n+1} = X_n + \gamma_{n+1} \left(h(X_n) + \Delta M_{n+1} \right), \tag{1.2}$$

où les termes $\Delta M_{n+1} := H(X_n, \epsilon_{n+1}) - h(X_n)$, $n \geq 0$ sont des accroissements de la martingale $(M_n)_{n\geq 1}$ par rapport à la filtration $\mathcal{F}_n := \sigma(X_0, \epsilon_1, \cdots, \epsilon_n)$.

L'étude des algorithmes stochastiques débuta dans les années 1950 avec les travaux de Robbins-Monro [76] et Kiefer et Wolfowitz [53]. Elle fût l'objet de nombreux travaux liés par exemple au contrôle adaptatif (c.f. [57]) ou aux estimations récursives (c.f. par exemple [69]), ... Dans le cadre de ce travail, nous nous intéressons aux questions de contrôle et de gestion du risque sur les marchés financiers et de l'énergie.

Ce premier chapitre de la thèse est découpé en trois parties.

Dans la première partie, nous étudions l'estimation de la VaR et de la CVaR en utilisant la théorie des algorithmes stochastiques. Cette première étude est réalisée dans le cas où la perte du portefeuille L s'écrit $\varphi(X)$ où φ est une fonction mesurable à valeurs dans $\mathbb{R}$ et X est un vecteur aléatoire structurel à valeurs dans $\mathbb{R}^d$. Partant de la caractérisation de la VaR et de la CVaR introduite dans [78], nous proposons un algorithme stochastique à deux composantes afin d'estimer récursivement ces deux

quantités. Des résultats classiques de la théorie des algorithmes stochastiques (c.f. [23], [22], [57] et [13]) nous permettent d'établir facilement la convergence presque sûre de l'algorithme proposée vers le couple VaR-CVaR. De la même manière, nous obtenons la vitesse de convergence faible de l'algorithme vers sa cible: elle est donnée par un théorème central limite (TCL) gaussien. Néanmoins la vitesse de convergence de l'algorithme est lente en pratique surtout lorsque le niveau de confiance α est proche de 1 (ou de 0). En effet, la VaR et la CVaR sont des quantités liées à la queue de distribution des pertes, leur estimation fait donc intervenir des évènements rares et doit donc être combinée avec une méthode de réduction de variance en pratique.

La plus adaptée à cette problématique est l'échantillonnage préférentiel (ou Importance Sampling). Pour cette raison, nous proposons de combiner notre première procédure avec un algorithme d'échantillonnage préférentiel récursif. L'algorithme stochastique ainsi obtenu vérifie un TCL avec une vitesse de convergence optimale. Cependant, lorsque la dimension d du vecteur X est grande, par exemple dans le cas où X est un vecteur d'accroissements browniens liés au schéma d'Euler d'une diffusion, des problèmes de convergence des paramètres de réduction de variance apparaissent.

Afin de remédier à ce problème, nous propons dans la deuxième partie de ce chapitre une extension de l'étude précédente. Nous nous intéressons au cas où la perte du portefeuille s'écrit $\varphi(X)$ où φ est à présent une fonctionnelle définie sur l'espace $\mathcal{C}\left([0,T],\mathbb{R}^d\right)$ des fonctions continues de $[0,T]$ à valeurs dans $\mathbb{R}^d$ et $X = (X_t)_{t\in[0,T]}$ est un processus d'Itô à valeurs dans $\mathbb{R}^d$ solution d'une équation différentielle stochastique. Afin de suppléer aux défauts de l'algorithme stochastique d'échantillonnage préférentiel en dimension finie, nous proposons une nouvelle procédure récursive et adaptative basée sur l'identité de Girsanov. Celle-ci se combine facilement avec la procédure d'estimation de la VaR et de la CVaR qui reste en tout point identique à sa première version en dimension finie. L'algorithme ainsi obtenu vérifie un TCL dont la vitesse de convergence est optimale. L'efficacité de la méthode est étudiée sur un ensemble de portefeuilles et une comparaison avec la procédure en dimension finie est proposée qui prouve l'utilité de notre approche.

Dans la première étude que nous avons menée, à chaque pas de l'algorithme VaR-CVaR, la génération pseudo-aléatoire du vecteur X à valeurs dans $\mathbb{R}^d$ est généralement obtenue comme une fonction d'un autre vecteur pseudo-aléatoire $U \sim \mathcal{U}\left([0,1]^s\right)$ $(s \geq d)$: $X = \Psi(U)$ en loi. Une idée naturelle est de remplacer à chaque pas de l'algorithme la simulation pseudo-aléatoire d'une copie de U, qui est déterministe mais dont le but est de reproduire les propriétés idéales des sources complètement aléatoires, *i.e.* les lois du hasard, par la valeur déterministe x_n d'une suite équirépartie $x = (x_n)_{n\geq 1}$ sur $[0,1]^s$ dont le but est de remplir l'espace $[0,1]^s$ de le plus uniformément possible. C'est pour cette raison que dans la troisième partie, nous nous intéressons à l'algorithme VaR-CVaR étudié dans la première partie de ce chapitre dans le cas où les innovations de l'algorithme ne sont plus obtenues via des nombres pseudo-aléatoires mais par une suite équirépartie déterministe à discrépance faible. Etant donnée l'efficacité des méthodes de Quasi-Monte Carlo appliquée au calcul d'espérance, l'idée d'utiliser des suites à discrépance faible dans le cadre des algorithmes stochastiques semble naturelle. Une première étude menée dans [58] montre que sous certaines hypothèses (d'une certaine manière plus restrictives que dans le cadre pseudo-aléatoire), ces algorithmes déterministes "quasi-stochastiques"

convergent et que leur vitesse de convergence est meilleure que celle de leur homologue stochastique. De nombreux résultats numériques appuient les conclusions de cette première étude. Cependant, ces résultats théoriques ne s'appliquent pas à notre algorithme. En utilisant un résultat récent obtenu dans [60] basé sur une hypothèse de moyennisation du processus d'innovations et des hypothèses de Lyapunov classiques, ainsi que des résultats sur la discrépance d'ensemble Jordan mesurable (c.f. [70] et [71]), nous démontrons la convergence de l'algorithme VaR-CVaR. Bien que dans ce cadre non stochastique, les techniques de réduction de variance en tant que telles n'ont théoriquement pas lieu d'être, nous montrons que d'un point de vue numérique l'algorithme préférentiel accélère la convergence de l'algorithme VaR-CVaR.

Dans le deuxième chapitre de cette thèse, nous nous intéressons à la couverture du risque dans un marché incomplet opérant à temps discret. Plus précisément, nous proposons une méthode de couverture d'une perte (ou d'un actif) dépendant d'une source de risque observable mais non négociable sur les marchés financiers ou de l'énergie (par exemple, la température). La stratégie autofinancée optimale est obtenue en minimisant dynamiquement la CVaR. Pour cela nous utilisons trois outils: les algorithmes stochastiques, la quantification vectorielle optimale et deux techniques de réduction de variance (l'échantillonnage préférentiel et les variables de contrôle). Dans un premier temps, nous introduisons la notion de CVaR dynamique qui est une version aléatoire et dynamique de la CVaR classique. Ensuite, nous étudions le cas simple mais intéressant des stratégies à un pas, *i.e.* à un rebalancement. Une telle stratégie décidée à la date $t_{\ell_0} \in [0,T]$ est une stratégie où le nombre d'actifs négociables est fixé à la date $t_{\ell_0} \in [0,T]$ et reste inchangé jusqu'à la date T. Nous établissons un résultat d'existence de stratégie optimale sous des hypothèses simples. En utilisant des idées similaires au principe de la programmation dynamique, nous établissons l'existence d'une stratégie optimale dans le cas des stratégies auto-financées dynamiques sous des hypothèses de non dégénérescence du processus de prix.

D'un point de vue numérique, nous nous plaçons sous l'hypothèse principale que le processus constitué par le couple risque observable mais non négociable et prix est markovien. Nous supposons également que le processus de prix est une martingale sous la filtration propre des deux processus. Nous proposons un algorithme stochastique afin d'estimer la stratégie optimale ainsi que la VaR et la CVaR associées au portefeuille avec couverture. Nous nous appuyons sur ce premier algorithme lorsque nous nous intéressons au cas des stratégies dynamiques. Nous proposons et comparons quatre stratégies dynamiques sous-optimales différentes afin d'approcher la stratégie optimale.

Ces résultats sont appliqués à la couverture de plusieurs portefeuilles sur les marchés de l'énergie, typiquement ceux de l'électricité et du gaz. Ces marchés sont incomplets pour plusieurs raisons: les sous-jacents comme l'électricité et le gaz ne sont pas stockables (ou difficilement et surtout pas à moindre coût), les prix de l'electricité et du gaz dépendent clairement de la température, ...

La troisième partie traite de la modélisation conjointe des prix spot du gaz et de l'électricité. Nous proposons un modèle fondé sur des processus de retour à la moyenne (processus d'Ornstein) markovien dont le coefficient de diffusion est paramétrique. Les paramètres du modèle sont choisis de façon à reproduire la struc-

ture de corrélation des deux énergies ainsi que l'ensemble des propriétés statistiques des prix spot: stationnarité, pics de prix, distributions à queues épaisses. La méthode de calibration que nous utilisons est basée sur des outils statistiques standards et robustes: méthodes des moindres carrés, méthode du maximum de vraisemblance (tronquée). Nous appliquons ce modèle au marché anglais du gaz et au marché français de l'électricité. Nous illustrons l'importance de la prise en compte de la structure de corrélation des prix spot du gaz et de l'électricité et de la présence des pics de prix en mesurant le risque sur un portefeuille lié au marché de l'énergie.

Dans la suite de cette introduction, nous allons exposer la problèmatique de chaque chapitre ainsi que les résultats importants obtenus.

1.1 Estimation de la VaR et de la CVaR par algorithme stochastique

Le contrôle et la gestion des risques est une composante de plus en plus importante de toutes activités de marché. Contrôler le risque c'est avant toute chose le mesurer à l'aide d'indicateurs pertinents et de procédures fiables. Le risque de marché est présent dès lors qu'une position sur les marchés financiers ou de l'énergie peut entraîner des pertes importantes. Les autorités de réglementation (Bâle II) prônent et exigent l'utilisation de la *Value-at-Risk* (VaR) comme mesure de risque. De la même manière, face à la croissance des marchés des matières premières et de l'énergie, les énergéticiens utilisent couramment la VaR pour mesurer le risque lié aux pertes de leurs divers portefeuilles d'actifs physiques ou financiers. Cependant, la VaR a suscité beaucoup de critiques émanant tant des practiciens que des milieux académiques. La VaR n'est pas une mesure cohérente du risque (selon la terminologie introduite dans [3]): la VaR de deux portefeuilles intégrés peut être supérieure à la somme des VaRs de chacun des portefeuilles, ce qui va à l'encontre du principe de diversification. De plus, la VaR ne dit rien sur les pertes effectives en cas de dépassement. Ces pertes peuvent être très élevées si les queues de distribution des pertes sont épaisses. Afin de palier les problèmes de la VaR, d'autres mesures de risque ont donc été proposées comme la *Conditional Value-at-Risk* (CVaR). Cette dernière est une mesure de risque cohérente. Elle représente la moyenne des pertes lorsque celles-ci dépassent le seuil de la VaR. Dès lors, la CVaR permet de mieux cerner ce qui peut se passer lorsque des évènements anormaux surviennent sur les marchés et de donner une estimation de la perte moyenne du portefeuille dans les mauvais jours. Mesurer efficacement le risque des pertes de portefeuilles lourds et complexes en actifs détenus est un défi de taille. Cela est la principale motivation qui nous a conduit à nous intéresser au problème d'estimation de la VaR et de la CVaR.

Ce problème est très largement étudié dans la littérature. D'un point de vue mathématique, sur un intervalle de temps donné $[t, T]$, la perte du portefeuille L est définie comme l'opposé de la différence de valeur de portefeuille entre la date t et la date T: $L := -\Delta V = V(S_t, t) - V(S_t + \Delta S, T)$, où S_t représente les différentes sources de risque du portefeuille au sein du marché considéré (prix de marché, taux, aléas climatiques, ...) observées à la date t, $\Delta S = S_T - S_t$ représente la variation de

S sur l'intervalle de temps $[t, T]$ et $V(S_t, t)$ représente la valeur du portefeuille à la date t. Contrairement à la valorisation et à la couverture des produits dérivés, où les prix de marché sont modélisés sous la probabilité risque neutre, la distribution des variations ΔS qui est pertinente est la distribution observée sous la probabilité historique. La VaR au seuil α est le plus petit quantile de L au seuil α, *i.e.*:

$$\mathrm{VaR}_\alpha(L) := \inf\{\xi \mid \mathbb{P}(L \leq \xi) \geq \alpha\}.$$

Nous travaillerons sous l'hypothèse que la distribution L est continue (*i.e.* sans atomes), ainsi la VaR_α est la plus petite solution de l'équation

$$\mathbb{P}(L \leq \xi) = \alpha.$$

Si la fonction de répartition de la perte L est (strictement) croissante alors la solution à cette équation est unique, sinon il peut y en avoir plusieurs. Si L vérifie $\mathbb{E}[L_+] < +\infty$, la CVaR_α est définie par

$$\mathrm{CVaR}_\alpha(L) := \mathbb{E}[L | L \geq \mathrm{VaR}_\alpha(L)].$$

De nombreuses méthodes ont été proposées pour estimer la VaR:

- La plus simple suppose que la distribution ΔS est une loi normale multivariée centrée $\mathcal{N}(0, \Sigma_S)$ (la matrice de variance covariance Σ_S est connue ou estimée) et que la perte est linéaire en ΔS:

$$L \approx -\delta^T \Delta S, \tag{1.3}$$

  où δ est le vecteur de sensibilité dont la ième composante est égale à $\delta_i = \frac{\partial V}{\partial S_i}(S_t, t)$, où V représente la valeur du portefeuille à la date t. Ce vecteur de sensibilité est la plupart du temps estimé par les entités de contrôle des risques. Alors la perte est de loi normale $L \sim \mathcal{N}(0; \sigma^2)$ avec $\sigma^2 = \delta^T \Sigma_S \delta$. La VaR est alors facilement calculable. Néanmoins, beaucoup de portefeuilles ne dépendent pas linéairement des sources de risques. Une telle approximation reste assez grossière. L'approximation Delta-Gamma tente de palier ce problème en ajoutant des termes d'ordre 2 à (1.3):

$$L \approx \frac{\partial V}{\partial t}\Delta t + \delta^T \Delta S + \frac{1}{2}\Delta S^T \Gamma \Delta S,$$

  où chaque terme de la matrice Γ, $\Gamma_{i,j} = \frac{\partial^2 V}{\partial S_i \partial S_j}(S_t, t)$ est supposé connu ou facile à estimer. Ensuite, il est possible d'en déduire une approximation de la fonction de répartition de la perte L et de trouver par inversion la VaR_α. Pour plus de détails, nous renvoyons à [17], [40], [39], [80] parmi d'autres. Cependant ces deux méthodes ne fonctionnent plus lorsque la maturité $T - t$ du portefeuille est grande ($T - t$ est de l'ordre de la semaine dans le secteur bancaire alors que $T - t \geq 1,2$ mois pour les marchés de l'énergie) ou lorsque L est une fonctionnelle d'un processus solution d'une EDS.

- La méthode par simulation historique ou Monte Carlo consiste dans un premier temps à générer n vecteurs de loi ΔS en utilisant des historiques de prix ou un modèle de prix et des variables pseudo-aléatoires. Puis, pour chaque vecteur ΔS simulé, on évalue la perte associée. Enfin, on inverse la fonction de répartition empirique de la perte L

$$\hat{F}_n(\xi) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{L_i \leq \xi\}},$$

et on obtient le quantile d'ordre α:

$$\xi_{n,\alpha} = \hat{F}_n^{-1}(\alpha).$$

Dans le cadre des simulations Monte Carlo, ce quantile converge presque sûrement vers la VaR_α lorsque celle-ci est unique. De plus, si L admet une densité f_L strictement positive au voisinage de $\xi_\alpha^* := \mathrm{VaR}_\alpha$, alors

$$\sqrt{n}\left(\xi_{n,\alpha} - \xi_\alpha^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0; \sigma_\alpha^2\right), \quad n \to +\infty, \tag{1.4}$$

où $\sigma_\alpha^2 = \frac{\alpha(1-\alpha)}{f_L^2(\xi_\alpha^*)}$. Pour une preuve de ce résultat, nous nous réferons à [84].

- Dans le cadre de la couverture ou plutôt de l'optimisation de portefeuille pour réduire la CVaR, R.T. Rockafellar et S. Uryasev dans [77] ont montré que la VaR et la CVaR sont respectivement la solution et la valeur d'un problème de minimisation convexe. Plus précisément,

Proposition 1.1.1. *Soit V la fonction définie par*

$$V(\xi) = \mathbb{E}\left[v(\xi, L)\right],$$

où

$$v(\xi, L) := \xi + \frac{1}{1-\alpha}\left(L - \xi\right)_+.$$

On suppose que la fonction de répartition de L est continue, croissante et que $\mathbb{E}\left[L_+\right] < +\infty$. Alors la fonction V est convexe, Lipschitz continue, dérivable, satisfait $\lim_{|\xi|\to+\infty} V(\xi) = +\infty$ et la VaR_α est définie par

$$\arg\min V = \{\xi \in \mathbb{R} \mid V'(\xi) = 0\} = \{\xi \mid \mathbb{P}(L \leq \xi) = \alpha\},$$

où V' est la dérivée de V définie pour tout $\xi \in \mathbb{R}$ par

$$V'(\xi) = \mathbb{E}\left[\frac{\partial v}{\partial \xi}(\xi, L)\right].$$

De plus,

$$\mathrm{CVaR}_\alpha(L) = \min_{\xi\in\mathbb{R}} V(\xi)$$

Dans la pratique, la plupart des institutions de contrôle des risques sont intéressées par le cas où α est proche de 1. Trois valeurs de α sont couramment utilisées: 95%, 99%, 99.5%. Par conséquent, la VaR_α se situe dans la queue de distribution extrême de droite. Dans cette situation, les évenements intéressants sont observés avec une probabilité très faible (moins de 5%). Par conséquent, ces méthodes doivent être combinées avec des procédures de réduction de variance. La plus adaptée au cadre des évenements rares est la méthode d'échantillonnage préférentiel (importance sampling). Appliquée à l'estimation de la VaR, [39] propose une méthode d'échantillonnage préférentiel basée sur une analyse par grande déviation qui fournit de bons résultats en terme de réduction de variance. Cependant, dans [41] ces résultats sont contrastés. En effet, les auteurs remarquent qu'une telle analyse peut conduire à une variance qui augmente avec α, voire une variance infinie dans certains cas. Dans [25], la VaR est estimée par la méthode d'inversion de la fonction de répartition pondérée. Afin d'estimer les paramètres optimaux d'échantillonnage préférentiel, celle-ci est combinée avec un algorithme stochastique contraint (*i.e.* projeté). Cet algorithme fut étudié dans le cadre gaussien dans [1]. La convergence s'effectue après une période de stabilisation si la suite de compacts sur lesquels la projection s'effectue a été spécifiée correctement.

Dans les trois parties de ce premier chapitre, nous proposons une nouvelle méthode d'estimation de la VaR et la CVaR par simulation basée sur les algorithmes stochastiques et une méthode d'échantillonnage préférentiel récursive non contrainte.

1.1.1 Cas de la dimension finie

Dans la première partie, nous étudions le cas où la perte L s'écrit: $L = \varphi(X)$ où $\varphi : \mathbb{R}^d \to \mathbb{R}$ est une fonction borélienne et X est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ modélisant ΔS sur l'horizon de temps considéré. Si on pose

$$H_1(\xi, x) = \frac{\partial v}{\partial \xi}(\xi, x) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(x)\geq\xi\}},$$

la proposition 1.1 montre que la $\mathrm{VaR}_\alpha := \xi_\alpha^*$ est une solution de l'équation $\mathbb{E}\left[H_1(\xi, X)\right] = 0$. Par conséquent, une méthode possible pour estimer ξ_α^* est d'implémenter l'algorithme de gradient stochastique:

$$\xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n),\ n \geq 1,$$

où $(X_n)_{n\geq 1}$ est une suite i.i.d. de variables aléatoires de même loi que X, indépendante de ξ_0, avec $\mathbb{E}\left[|\xi_0|^2\right] < +\infty$ et $(\gamma_n)_{n\geq 1}$ est une suite de pas déterministe et positive vérifiant

$$\sum_{n\geq 1}\gamma_n = +\infty \quad \text{et} \quad \sum_{n\geq 1}\gamma_n^2 < +\infty. \tag{1.5}$$

Sous des hypothèses appropriées, nous montrons à l'aide du théorème classique de Robbins-Monro que $\xi_n \xrightarrow{p.s.} \xi_\alpha^*$, $n \to +\infty$. Pour estimer la $\mathrm{CVaR}_\alpha := C_\alpha^*$ nous proposons la *procédure compagnon*,

$$C_n = C_{n-1} - \gamma_n H_2\left(\xi_{n-1}, C_{n-1}, X_n\right),\ n \geq 1,$$

où $H_2(\xi, c, x) := c - v(\xi, x)$. Dès que la distribution des pertes satisfait $\varphi(X) \in L^2(\mathbb{P})$ et que (1.5) est vérifiée, nous obtenons $(\xi_n, C_n) \xrightarrow{p.s.} (\xi_\alpha^*, C_\alpha^*)$.

La vitesse de convergence faible de la procédure $(\xi_n, C_n)_{n\geq 1}$ est régie par un TCL classique (c.f. les livres [23], [13] ou [57] par exemple). La vitesse de convergence optimale est obtenue en choisissant un pas $\gamma_n = \frac{a}{b+n}$, $a, b > 0$. Cependant, le choix de la constante a est sujet à une condition faisant intervenir $f_L(\xi_\alpha^*)$. En pratique, cela nous conduit à choisir a de manière arbitraire. Afin de résoudre ce problème, une méthode couramment utilisée est le principe de moyennisation de Ruppert & Polyak (c.f. les deux articles fondateurs [81] et [45]). On écrit tout d'abord l'algorithme VaR-CVaR de manière plus synthétique pour $n \geq 1$

$$\phi_n = (\xi_n, C_n),\ \phi_0 = (\xi_0, C_0),$$

et,

$$\phi_n = \phi_{n-1} - \gamma_n H(\phi_{n-1}, X_n),$$

où $H(\phi, x) = (H_1(\xi, x), H_2(\xi, c, x))$. Ensuite, nous calculons de façon adaptative la moyenne de Cesàro de la procédure,

$$\overline{\phi}_n = \frac{\phi_0 + \cdots + \phi_{n-1}}{n}, \quad n \geq 1.$$

La nouvelle suite $\left(\overline{\phi}_n\right)_{n\geq 1}$ converge presque sûrement vers $\phi^* = (\xi_\alpha^*, C_\alpha^*)$. Un choix approprié de la suite de pas $(\gamma_n)_{n\geq 1}$ (pas lentement décroissant) garantit la convergence de $\left(\overline{\phi}_n\right)_{n\geq 1}$ à la vitesse optimale $\sqrt{n}$ avec une variance asymptotique minimale.

Theorem 1.1.2. *On suppose que la distribution des pertes vérifie $\varphi(X) \in L^{2a}(\mathbb{P})$ pour un $a > 1$ et que la densité de L au point ξ_α^* est strictement positive. Si la séquence de pas est définie par $\gamma_n = \frac{\gamma_1}{n^\beta}$ avec $\frac{1}{2} < \beta < 1$ et $\gamma_1 > 0$ alors*

$$\sqrt{n}\left(\overline{\phi}_n - \phi^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

où la matrice de variance covariance asymptotique Σ est définie par

$$\begin{pmatrix} \frac{\alpha(1-\alpha)}{f_L^2(\xi^*)} & \frac{\alpha}{(1-\alpha)f_L(\xi^*)}\mathbb{E}\left[(\varphi(X) - \xi^*)_+\right] \\ \frac{\alpha}{(1-\alpha)f_L(\xi^*)}\mathbb{E}\left[(\varphi(X) - \xi^*)_+\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\varphi(X) - \xi^*)_+\right) \end{pmatrix}. \tag{1.6}$$

D'un point de vue numérique, l'algorithme VaR-CVaR converge lentement. En effet, le désavantage de cette première version réside dans la mise à jour de l'estimateur de la VaR qui n'est effectuée que lorsqu'un évènement rare du type $\varphi(X) > \xi_\alpha^*$ intervient. Celui-ci arrive avec une probabilité très faible: $\mathbb{P}(\varphi(X) > \xi_\alpha^*) = 1 - \alpha \approx 0$ lorsque α est proche de 1, autrement la procédure converge rapidement. De plus, dans la pratique, nous sommes souvent limités en nombre de scénarios à cause de la complexité des portfeuilles utilisés. Pour ces deux raisons, il est nécessaire de combiner notre premier algorithme avec une méthode de réduction de variance. Une méthode adaptée à notre problématique est l'échantillonnage préférentiel. Celle-ci consiste à modifier la distribution de la perte L afin de donner plus de poids aux scénarios "importants". Nous supposons que X admet une densité p.

Appliqué au calcul de l'espérance $\mathbb{E}[F(X)]$, l'échantillonnage préférentiel par translation consiste à utiliser l'invariance par translation de la mesure de Lebesgue sur $\mathbb{R}^d$ pour introduire un paramètre $\theta \in \mathbb{R}^d$

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X+\theta)\frac{p(X+\theta)}{p(X)}\right].$$

Dès que $F(X) \in L^2(\mathbb{P})$ satisfait $\mathbb{P}(F(X) \neq 0) > 0$, on désire choisir parmi toutes ces nouvelles variables aléatoires de même espérance celle qui minimise la variance, *i.e.* le moment d'ordre 2

$$Q(\theta) := \mathbb{E}\left[F^2(X+\theta)\frac{p^2(X+\theta)}{p^2(X)}\right] = \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right] \leq +\infty, \quad \theta \in \mathbb{R}^d.$$

Plusieurs procédures ont été proposées dans la littérature afin d'estimer le paramètre optimal $\theta^* := \operatorname{Arg\,min} Q$. Lorsque X est un vecteur gaussien, B. Arouna (c.f. [1]) propose un algorithme stochastique projeté (Projection à la Chen) basé sur la représentation de ∇Q sous forme d'espérance. V. Lemaire & G. Pagès dans [62] ont proposé récemment un nouvel algorithme stochastique non-projeté basé sur une nouvelle représentation de ∇Q sous forme d'espérance $\nabla Q(\theta) = \mathbb{E}[K(\theta, X)]$ et sur un contrôle de F à l'infini,

$$\theta_n = \theta_{n-1} - \gamma_n \Psi(\theta_n) K(\theta_{n-1}, X_n),\ \theta_0 \in \mathbb{R}^d,$$

où $\Psi(\theta) > 0$, pour tout $\theta \in \mathbb{R}^d$ est une fonction permettant de "contrôler" le comportement de F à l'infini. Sous certaines hypothèses de contrôle de F à l'infini et sur la densité p, ils obtiennent $\theta_n \xrightarrow{p.s.} \theta^*$. C'est cette dernière version que nous allons utiliser pour réduire la variance de notre algorithme VaR-CVaR. Dans notre cas, il s'agit de minimiser les deux variances asymptotiques

$$\frac{\alpha(1-\alpha)}{f_L^2(\xi_\alpha^*)} = \frac{\mathrm{Var}(\mathbf{1}_{\varphi(X)\geq\xi^*})}{f_L^2(\xi_\alpha^*)} \quad \text{pour la } \mathrm{VaR}_\alpha, \tag{1.7}$$

et,

$$\frac{\mathrm{Var}((\varphi(X)-\xi^*)_+)}{(1-\alpha)^2} \quad \text{pour la } \mathrm{CVaR}_\alpha, \tag{1.8}$$

donc de minimiser les deux moments d'ordre 2,

$$Q_1(\theta, \xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p(X-\theta)}\right],$$
$$Q_2(\mu, \xi^*) := \mathbb{E}\left[(\varphi(X)-\xi^*)_+^2\,\frac{p(X)}{p(X-\mu)}\right].$$

Ainsi, nous obtenons deux procédures récursives

$$\theta_n := \theta_{n-1} - \gamma_n K_1(\xi_{n-1}, \theta_{n-1}, X_n),$$
$$\mu_n := \mu_{n-1} - \gamma_n K_2(\xi_{n-1}, \mu_{n-1}, X_n),$$

que nous combinons adaptativement (*i.e.* en utilisant les mêmes innovations) avec notre premier algorithme VaR-CVaR:

$$\xi_n := \xi_{n-1} - \gamma_n L_1(\xi_{n-1}, \theta_{n-1}, X_n),$$
$$C_n := C_{n-1} - \gamma_n L_2(\xi_{n-1}, \mu_{n-1}, X_n),$$

où

$$L_1(\xi,\theta,x) := e^{-\rho|\theta|^b}\left(1-\frac{1}{1-\alpha}\ \mathbf{1}_{\{\varphi(x+\theta)\geq\xi\}}\frac{p(x+\theta)}{p(x)}\right),$$
$$L_2(\xi,C,\mu,x) := C-\xi-\frac{1}{1-\alpha}(\varphi(x+\mu)-\xi)_+\frac{p(x+\mu)}{p(x)},$$

avec $\rho>0$ et $b\in[1,2]$. Nous montrons sous certaines hypothèses $(\xi_n,C_n,\theta_n,\mu_n)\xrightarrow{p.s.}(\xi^*_\alpha,C^*_\alpha,\theta^*_\alpha,\mu^*_\alpha)$, $n\to+\infty$. Quant à la vitesse de convergence faible, la procédure obtenue vérifie, sous certaines hypothèses que nous énoncerons plus tard, un TCL gaussien.

Theorem 1.1.3. *On suppose que la séquence des pas $(\gamma_n)_{n\geq1}$ est définie pour $n\geq1$ par $\gamma_n=\frac{\gamma_1}{n^\beta}$, avec $\frac{1}{2}<\beta<1$ et que les hypothèses du théorème 1.1.2. sont satisfaites. La séquence $(\overline{\xi}_n,\overline{C}_n)_{n\geq1}$ définie par*

$$\overline{\xi}_n := \frac{\xi_0+...+\xi_{n-1}}{n},\quad \overline{C}_n := \frac{C_0+...+C_{n-1}}{n},\quad n\geq1,$$

satisfait

$$\sqrt{n}\begin{pmatrix}\overline{\xi}_n-\xi^*\\ \overline{C}_n-C^*\end{pmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Sigma^*)\quad,\ n\to+\infty, \tag{1.9}$$

avec

$$\begin{aligned}
\Sigma^*_{1,1} &= \frac{1}{f_L^2(\xi^*)}\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)\geq\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right),\\
\Sigma^*_{1,2} &= \Sigma^*_{2,1}=\frac{1}{(1-\alpha)f_L(\xi^*)}\mathrm{Cov}\left((\varphi(X+\mu^*_\alpha)-\xi^*)_+\frac{p(X+\mu^*_\alpha)}{p(X)},\right.\\
&\qquad\left.\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)\geq\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right),\\
\Sigma^*_{2,2} &= \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\varphi(X+\mu^*_\alpha)-\xi^*)_+\frac{p(X+\mu^*_\alpha)}{p(X)}\right).
\end{aligned}$$

Un paragraphe de cette partie est consacré à l'application de cet algorithme d'estimation de la VaR et de la CVaR à de nombreux portefeuilles typiques rencontrés sur le marché de l'énergie. Les résultats numériques montrent que la réduction de variance est très significative. Elle l'est d'autant plus que α est proche de 1, *i.e.* plus on cherche à simuler des évènements extrêmes plus la méthode d'échantillonnage préférentiel est efficace.

Par la même occasion, nous proposons d'autres algorithmes de réduction de variance lorsque la méthode d'échantillonnage préférentiel s'appuie sur la transformation d'Esscher (changement de mesure exponentiel). Cette dernière transformation est très efficace lorsque le vecteur X est à queue lourde comme c'est le cas pour les distributions hyperboliques généralisées.

Les estimateurs des paramètres d'échantillonnage préférentiel $(\theta_n,\mu_n)_{n\geq1}$ sont chacun de dimension d. Lorsque celle-ci devient grande ($d\geq50$ en pratique) comme c'est souvent le cas des portefeuilles des praticiens, où par exemple le vecteur X représente les accroissements browniens liés au schéma d'Euler d'une diffusion, l'algorithme récursif d'échantillonnage préférentiel présente des difficultés de convergence et n'est plus utilisable.

1.1.2 Cas de la dimension infinie

Dans la deuxième partie, nous étudions le cas où la perte L s'écrit: $L = \varphi(X)$ où φ est une fonctionnelle définie sur l'espace $\mathcal{C}\left([0,T],\mathbb{R}^d\right)$ des fonctions continues de $[0,T]$ dans $\mathbb{R}^d$ et la source de risque X est un processus d'Itô solution d'une EDS:

$$\mathrm{d}X_t = b(t,X^t)\mathrm{d}t + \sigma(t,X^t)\mathrm{d}W_t, \quad X_0 = x \in \mathbb{R}^d, \tag{$E_{b,\sigma}$}$$

où $W = (W_t)_{t\in[0,T]}$ est un mouvement brownien de dimension q, $X^t := (X_{t\wedge s})_{s\in[0,T]}$ est le processus arrêté au temps t et $b : [0,T]\times\mathcal{C}\left([0,T],\mathbb{R}^d\right) \to \mathbb{R}^d$, $\sigma[0,T]\times\mathcal{C}\left([0,T],\mathbb{R}^d\right) \to \mathcal{M}(d,q)$ sont des fonctions mesurables vérifiant les deux conditions suivantes:

(i) $b(.,0)$ et $\sigma(.,0)$ sont continues,
(ii) $\forall t\in[0,T], \forall x,y \in \mathcal{C}\left([0,T],\mathbb{R}^d\right), |b(t,y)-b(t,x)|$
$$+ \|\sigma(t,y)-\sigma(t,x)\| \le C_{b,\sigma}\|x-y\|_\infty .$$

L'algorithme VaR-CVaR sans méthode de réduction de variance reste identique à celui développé dans la première partie de ce chapitre. Cependant, l'algorithme d'échantillonnage préférentiel est maintenant fondé sur la transformation de Girsanov. Si f est une fonction borélienne bornée de $\mathcal{C}\left([0,T],\mathbb{R}^d\right)$ à valeurs dans $\mathcal{M}(d,q)$ et si $\theta \in L^2_{T,p} := L^2([0,T],\mathbb{R}^p)$, la transformation de Girsanov appliquée au calcul de $\mathbb{E}[F(X)]$ s'écrit

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X^{(\theta)})e^{-\int_0^T\left\langle f(X^{(\theta),s})\theta_s,\mathrm{d}W_s\right\rangle - \frac{1}{2}\left\|f(X^{(\theta),\cdot})\theta_\cdot\right\|^2_{L^2_{T,q}}}\right],$$

où $X^{(\theta)}$ est la solution de $(E_{b+\sigma f\theta,\sigma})$. Dès que $F(X)\in L^2(\mathbb{P})$ satisfait $\mathbb{P}(F(X)\neq 0) > 0$, on désire choisir parmi toutes ces nouvelles variables aléatoires de même espérance, celle qui minimise la variance, *i.e.* le moment d'ordre 2

$$\begin{aligned} Q(\theta) &= \mathbb{E}\left[F(X^{(\theta)})^2 e^{-2\int_0^T\left\langle f(X^{(\theta),s})\theta_s,\mathrm{d}W_s\right\rangle - \left\|f(X^{(\theta),\cdot})\theta_\cdot\right\|^2_{L^2_{T,q}}}\right] \\ &= \mathbb{E}\left[F(X)^2 e^{-\int_0^T\left\langle f(X^{(\theta),s})\theta_s,\mathrm{d}W_s\right\rangle + \frac{1}{2}\left\|f(X^{(\theta),\cdot})\theta_\cdot\right\|^2_{L^2_{T,q}}}\right]. \end{aligned}$$

En pratique, nous minimisons Q sur un sous-espace fini $E = \mathrm{Vect}(e_1,\cdots,e_m) \subset L^2_{T,p}$. Dès que $\mathbb{E}[F(X)^{2+\eta}] < +\infty$ pour un $\eta > 0$ et que les deux conditions (i), (ii) sont vérifiées alors on peut montrer que Q est finie sur $L^2_{T,p}$, log-convexe, différentiable et $\lim_{\|\theta\|_{L^2_{T,p}}\to+\infty} Q(\theta) = +\infty$ ainsi $\mathrm{Arg}\min_\theta Q = \left\{\theta\in L^2_{T,p} \mid DQ(\theta) = 0\right\}$ est

non vide, la différentielle $DQ(\theta) \in L^2_{T,p}$ est donnée par

$$\begin{aligned}\langle DQ(\theta), \zeta\rangle_{L^2_{T,p}} &= \mathbb{E}\left[F(X)^2 e^{-\int_0^T \left\langle f(X^{(\theta),s})\theta_s, \mathrm{d}W_s\right\rangle + \frac{1}{2}\left\|f(X^{(\theta),\cdot})\theta_\cdot\right\|^2_{L^2_{T,q}}} \times \right. \\ &\qquad \left. \left(\langle f(X^\cdot)\theta_\cdot, f(X^\cdot)\zeta_\cdot\rangle_{L^2_{T,p}} - \int_0^T \langle f(X^s)\zeta_s, \mathrm{d}W_s\rangle\right)\right], \\ &= \mathbb{E}\left[F(X^{(-\theta)})^2 e^{\left\|f(X^{(\theta),\cdot})\theta_\cdot\right\|^2_{L^2_{T,q}}} \times \right. \\ &\qquad \left. \left(2\langle f(X^\cdot)\theta_\cdot, f(X^\cdot)\zeta_\cdot\rangle_{L^2_{T,p}} - \int_0^T \left\langle f(X^{(-\theta),s})\zeta_s, \mathrm{d}W_s\right\rangle\right)\right],\end{aligned}$$

où $\zeta \in L^2_{T,p}$. Pour une preuve de ce résultat, nous renvoyons à [62]. En ce qui concerne l'implémentation numérique, nous considérons un sous-espace E de $L^2_{T,p}$ non trivial de dimension finie. La restriction de Q sur E atteint un minimum θ^*_E. On considère alors $(e_1, \cdots, e_m)$ une base orthornormale de E. Sous des hypothèses de contrôle de la discrépance entre X et $X^{(-\theta)}$ et de contrôle de F à l'infini, on définit l'algorithme d'échantillonnage préférentiel par le schéma récursif

$$\theta_n = \theta_{n-1} - \gamma_n K\left(\theta_{n-1}, X^{(-\theta_{n-1})}, W^{(n)}\right),$$

où $(\gamma_{n\geq 1})_{n\geq 1}$ est une suite de pas vérifiant (1.5), $\left(W^{(n)}\right)_{n\geq 1}$ est une suite indépendante de mouvements browniens pour lesquels le processus $X^{(-\theta_{n-1})}$ est une solution forte de $E_{b-\sigma f(X^{(-\theta_{n-1})})\theta_{n-1}, W^{(n)}}$. La fonction K est définie sur la base $(e_1, \cdots, e_m)$ pour $\theta \in L^2_{T,p}$, W mouvement brownien, $Y = (Y_t)_{t\in[0,T]}$ processus $\mathcal{F}^W_t$-adapté à valeurs dans $\mathbb{R}^d$ par

$$\begin{aligned}\langle K(\theta, Y, W), e_i\rangle_{L^2_{T,p}} = C(\theta, f)F(Y)^2 e^{\|f(Y^\cdot)\theta_\cdot\|_{L^2_{T,q}}} &\left(2\langle f(Y^\cdot)\theta_\cdot, f(Y^\cdot)e_i\rangle_{L^2_{T,p}} \right. \\ &\left. - \int_0^T \langle f(Y^s)\zeta_s, \mathrm{d}W_s\rangle\right),\end{aligned}$$

où $C(\theta, f)$ est une constante positive dépendant de θ et f. La séquence $(\theta_n)_{n\geq 1}$ ainsi définie vérifie: $\theta_n \xrightarrow{p.s.} \theta^*_E$, $n \to +\infty$. Pour une preuve et plus de détails à propos de cet algorithme, nous nous référons à [62]. Dans notre cas, il s'agit toujours de minimiser les variances asymptotiques (1.7) et (1.8). Afin de définir notre algorithme d'échantillonnage préférentiel, nous supposons que φ vérifie l'hypothèse suivante:

$$\exists \lambda > 0,\ \forall x \in \mathcal{C}\left([0,T], \mathbb{R}^d\right),\quad |\varphi(x)| \leq C\left(1 + \|x\|^\lambda_\infty\right). \tag{1.10}$$

Nous obtenons ainsi deux procédures récursives

$$\begin{aligned}\theta_n &= \theta_{n-1} - \gamma_n K_1\left(\xi_{n-1}, \theta_{n-1}, X^{(-\theta_{n-1})}, W^{(n)}\right), \\ \mu_n &= \mu_{n-1} - \gamma_n K_2\left(\xi_{n-1}, \mu_{n-1}, X^{(-\mu_{n-1})}, W^{(n)}\right),\end{aligned}$$

que nous combinons adaptativement avec notre premier algorithme VaR-CVaR:

$$\begin{aligned}\xi_n &:= \xi_{n-1} - \gamma_n L_1(\xi_{n-1}, \theta_{n-1}, W^{(n)}), \\ C_n &:= C_{n-1} - \gamma_n L_2(\xi_{n-1}, \mu_{n-1}, W^{(n)}),\end{aligned}$$

où

$$L_1(\xi,\theta,W) := e^{-\frac{1}{2}\|f_1\|_\infty \|\theta_\cdot\|^2_{L^2_{T,p}}} \left(1 - \frac{1}{1-\alpha}\,\mathbf{1}_{\{\varphi(X^{(\theta)})\geq\xi\}} \times e^{-\int_0^T \left\langle f_1(X^{(\theta),s})\theta_s, \mathrm{d}W_s\right\rangle - \frac{1}{2}\left\|f_1(X^{(\theta),\cdot})\theta_\cdot\right\|^2_{L^2_{T,q}}}\right),$$

$$L_2(\xi,C,\mu,W) := C - \xi - \frac{1}{1-\alpha}(\varphi(X^{(\mu)}) - \xi)_+ e^{-\int_0^T \left\langle f_2(X^{(\mu),s})\mu_s, \mathrm{d}W_s\right\rangle - \frac{1}{2}\left\|f_2(X^{(\mu),\cdot})\mu_\cdot\right\|^2_{L^2_{T,q}}}.$$

Nous montrons que si les conditions (i) et (ii) sont vérifiées, $\varphi(X) \in L^2(\mathbb{P})$ et que l'hypothèse (1.10) est satisfaite alors $(\xi_n, C_n, \theta_n, \mu_n) \xrightarrow{p.s.} \left(\xi^*_\alpha, C^*_\alpha, \theta^*_{\alpha,E}, \mu^*_{\alpha,E}\right)$, $n \to +\infty$.

Quant à la vitesse de convergence faible, la procédure moyennisée (toujours en utilisant le principe de moyennisation de Ruppert & Polyak) vérifie sous certaines hypothèses un TCL gaussien.

Theorem 1.1.4. *On suppose que la suite de pas $(\gamma_n)_{n\geq 1}$ est définie pour $n \geq 1$ par $\gamma_n = \frac{\gamma_1}{n^\beta}$ avec $\frac{1}{2} < \beta < 1$. Si $\varphi(X) \in L^{2a}(\mathbb{P})$ pour un $a > 1$, (i) et (ii) sont vérifiées et si (1.10) est satisfaite alors la séquence moyennisée définie par*

$$\overline{\xi}_n := \frac{\xi_0 + ... + \xi_{n-1}}{n}, \quad \overline{C}_n := \frac{C_0 + ... + C_{n-1}}{n}, \quad n \geq 1,$$

satisfait

$$\sqrt{n}\begin{pmatrix} \overline{\xi}_n - \xi^*_\alpha \\ \overline{C}_n - C^*_\alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma^*) \quad , n \to +\infty, \tag{1.11}$$

avec

$$\Sigma^*_{1,1} = \frac{1}{f^2_L(\xi^*_\alpha)}\mathrm{Var}\left(\mathbf{1}_{\left\{\varphi\left(X^{(\theta^*_{\alpha,E})}\right)\geq\xi^*\right\}} \times e^{-\int_0^T \left\langle f_1(X^{(\theta^*\alpha,E),s})\theta^*_{\alpha,E,s}, dW_s\right\rangle - \frac{1}{2}\left\|f_1(X^{(\theta^*_{\alpha,E})^{,\cdot}})\theta^*_{\alpha,E,\cdot}\right\|^2_{L^2_{T,q}}}\right),$$

$$\Sigma^*_{2,2} = \frac{1}{(1-\alpha)^2}\mathrm{Var}\left(\left(\varphi\left(X^{(\mu^*_{\alpha,E})}\right) - \xi^*_\alpha\right)_+ \times e^{-\int_0^T \left\langle f_2(X^{(\mu^*\alpha,E),s})\mu^*_{\alpha,E,s}, dW_s\right\rangle - \frac{1}{2}\left\|f_2(X^{(\mu^*_{\alpha,E})^{,\cdot}})\mu^*_{\alpha,E,\cdot}\right\|^2_{L^2_{T,q}}}\right).$$

Un paragraphe de cette partie est consacré aux applications numériques sur de nombreux portefeuilles liés aux marchés de l'énergie. Les résultats numériques montrent que la réduction de variance est très significative. Notamment, nous comparons sur un même portefeuille l'efficacité de cette nouvelle version de l'algorithme avec sa version en dimension finie développée dans la partie précédente: l'algorithme VaR-CVaR en dimension infinie résout le problème de la dimension rencontré par l'algorithme VaR-CVaR en dimension finie.

1.1.3 Version QMC de l'approximation stochastique

Dans cette partie, nous étudions le cas où le vecteur aléatoire X n'est plus généré à l'aide d'une séquence (déterministe) pseudo-aléatoire mais à l'aide d'une suite (déterministe) uniformément distribuée à discrépance faible. Généralement, une variable aléatoire X de dimension d est simulée à partir d'une loi uniforme $\mathcal{U}([0,1]^q)$, avec $q \geq d$, par des méthodes standards tels que l'inverse de la fonction de distribution, la méthode de Box-Müller, ... Etant donné les performances des méthodes dites de quasi-Monte Carlo concernant l'intégration numérique, il paraît naturel d'essayer d'introduire des suites à discrépance faible dans les algorithmes stochastiques. La première étude fût menée par [58] où il est clairement montré (dans un cadre unidimensionel) que dans certains cas la convergence des algorithmes ainsi obtenus est bien meilleure que celle où une séquence pseudo-aléatoire est utilisée. D'un point de vue théorique, les deux principaux résultats sont basés sur une hypothèse de contraction et sur une hypothèse de bornitude de la fonction H. Dans le premier cas, une vitesse de convergence de l'algorithme vers sa cible est obtenue justifiant ainsi le gain par rapport à l'approximation stochastique classique basée sur des nombres pseudo-aléatoires.

Récemment, de nouveaux résultats concernant les algorithmes quasi-stochastiques ont été obtenus dans [60] (incluant d'autres cadres d'études) généralisant les résultats de [58]. Le résultat principal est basé sur une hypothèse de moyennisation de la séquence aléatoire utilisée dans l'algorithme stochastique et sur des hypothèses de Lyapunov classiques (retour à la moyenne et croissance linéaire). Cependant, ces résultats ne s'appliquent pas directement à l'algorithme VaR-CVaR car dans la majorité des cas, l'hypothèse de moyennisation n'est pas vérifiée. Dans cette partie, nous présentons un cadre dans lequel nous obtenons la convergence de l'algorithme VaR-CVaR. L'idée principale est d'utiliser des résultats concernant la discrépance pour des suites uniformément distribuées à valeurs dans des ensembles Jordan mesurable (c.f. [70] et [71]). Cela nous permet de satisfaire une hypothèse de moyennisation.

Definition 1.1.1. Une séquence $(u_n)_{n\geq 1}$ à valeur sur $[0,1]^q$ est uniformément distribuée dans $[0,1]^q$ si

$$\frac{1}{n}\sum_{k=1}^{n}\delta_{u_k} \overset{(\mathbb{R}^q)}{\Longrightarrow} \mathcal{U}([0,1]^q), \quad n \to +\infty,$$

où δ_u est la masse de Dirac en u, $\overset{(\mathbb{R}^q)}{\Longrightarrow}$ représente la convergence faible des mesures de probabilités sur $(\mathbb{R}^q, \mathcal{B}or(\mathbb{R}^q))$ et $\mathcal{U}([0,1]^q)$ représente la loi uniforme sur $[0,1]^q$.

La notion de discrépance permet de quantifier l'erreur de répartition des suites uniformément distribuées.

Proposition 1.1.5. *Une suite $(u_n)_{n\geq 1}$ à valeur dans $[0,1]^q$ est uniformément distribuée si et seulement si*

$$D_n^*(u) := \sup_{x\in[0,1]^q}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{[0,x]}(u_k) - \prod_{i=1}^{q}x_i\right| \to 0, \quad n \to +\infty.$$

$D_n^*(u)$ est la discrépance étoile.

On considère couramment qu'une suite uniformément distribuée sur $[0,1]^q$ est à discrépance faible si elle vérifie $D_n^*(u) = O\left(N^{-1}(\log N)^q\right)$.

L'idée principale de cette partie est fondée sur la notion de discrépance des sous-ensembles de $[0,1]^q$ qui sont Jordan-mesurables (*i.e.* les sous-ensembles dont la frontière est régulière ou dont la fonction caractéristique est intégrable au sens de Riemann). Cette discrépance sera très utile lorsque nous nous intéresserons à l'algorithme VaR-CVaR dans le cas des suites à discrépance faible. Nous nous réferons à [71] pour une preuve.

Proposition 1.1.6. *Soit $B \subseteq [0,1]^q$ et $\epsilon > 0$. Soient*

$$B_\epsilon = \{u \in [0,1]^q : \ d(u,v) < \epsilon \ \text{ pour un } v \in B\},$$
$$B_{-\epsilon} = \{u \in [0,1]^q : \ d(u,v) \geq \epsilon \ \text{ pour tout } v \in [0,1]^q \backslash B\},$$

où d est la distance Euclidienne sur $\mathbb{R}^q$. Soit $\mathcal{M}_b$ la famille d'ensemble $B \subseteq [0,1]^q$ mesurable au sens de Lebesgue vérifiant

$$\lambda_q\left(B_\epsilon \backslash B\right) \leq b(\epsilon) \quad \text{and} \quad \lambda_q\left(B \backslash B_{-\epsilon}\right) \leq b(\epsilon),$$

pour tout $\epsilon > 0$, où λ_q désigne la mesure de Lebesgue sur $(\mathbb{R}^q, \mathcal{B}or(\mathbb{R}^q))$ et b est une fonction positive, croissante tel que $\lim_{\epsilon \to 0^+} b(\epsilon) = 0$. Si $(u_n)_{n \geq 1}$ est uniformément distribuée sur $[0,1]^q$ alors

$$D_n(\mathcal{M}_b, u) := \sup_{B \in \mathcal{M}_b} \left| \frac{1}{n} \sum_{k=1}^{n} \delta_B(u_k) - \lambda_q(B) \right| \to 0, \quad n \to +\infty.$$

De plus, si la fonction b est de la forme $b(\epsilon) = C\epsilon$, pour tout $\epsilon > 0$ et pour une constante $C > 0$ alors il existe $K > 0$ tel que

$$D_n(\mathcal{M}_b, u) \leq K D_n^*(u)^{\frac{1}{q}}.$$

Les résultats suivants établissent un lien entre la discrépance et l'erreur numérique d'intégration d'une fonction f sur $[0,1]^q$. Ils fournissent l'erreur que l'on commet lorsqu'on estime la quantité $\mathbb{E}\left[f(U)\right]$, où $U \sim \mathcal{U}\left([0,1]^q\right)$ et f est une fonction à variation finie (au sens de Hardy et Krause) ou lipschitzienne, par la moyenne empirique $\frac{1}{n}\sum_{k=1}^n f(u_k)$.

Proposition 1.1.7. *Soit $u = (u_k)_{1 \leq k \leq n}$ une suite à valeurs dans $[0,1]^q$ et f une fonction à variation finie $V(f)$. Alors*

- *(Inégalité de Koksma-Hlwaka)*

$$\left| \frac{1}{n} \sum_{k=1}^{n} f(u_k) - \int_{[0,1]^q} f(u) \lambda_q(du) \right| \leq V(f) D_n^*(u).$$

- *([19]) Si $B \in \mathcal{M}_b$, alors on a*

$$\left| \frac{1}{n} \sum_{k=1, \ u_k \in B}^{n} f(u_k) - \int_B f(u) \lambda_q(du) \right| \leq \left(V(f) + f(1, \cdots, 1)\right) D_n(\mathcal{M}_b, u).$$

Un autre résultat très utile concerne le cas où la fonction f est lipschitzienne. Le théorème suivant est dû à Proinov (c.f. [75])

Theorem 1.1.8. *On suppose que $\mathbb{R}^q$ est muni de la norme ℓ^∞, ($|x|_\infty := \max_{1\leq i\leq q}|x_i|,\ x\in\mathbb{R}^q$). Soit u une suite uniformément distribuée à valeurs dans $[0,1]^q$. Soit $p_1,\cdots,p_n$ une suite de réels positifs vérifiant*

$$\sum_{k=1}^{n} p_k = 1.$$

Alors, pour toute fonction continue $f:[0,1]^q\to\mathbb{R}$,

$$\left|\sum_{k=1}^{n} p_k f(u_k) - \int_{[0,1]^q} f(x)\lambda_q(dx)\right| \leq C_q w_f\left(D_n^*(u)^{\frac{1}{q}}\right),$$

où

$$w_f(\delta) := \sup_{x,y\in[0,1]^q, |x-y|_\infty\leq\delta} |f(x)-f(y)|,\quad \delta\in(0,1),$$

et $C_q\in(0,\infty)$ est une constante universelle dépendant uniquement de q. Si $q=1$, $C_q=1$ et si $q\geq 2$, $C_q\in[1,4]$.

Pour revenir à l'algorithme VaR-CVaR, nous supposons que $X=\Psi(U)$, avec $\Psi:[0,1]^q\to\mathbb{R}^d$. Nous allons remplaçer U par une suite à discrépance faible u. Nous nous plaçons sous l'hypothèse suivante

Hypothèse 1.1.9. *La fonction $\varphi\circ\Psi:[0,1]^q\to\mathbb{R}$ est lipschitz.*

Soit F la fonction de distribution de la perte L, *i.e.* la fonction de distribution de la variable aléatoire $\varphi(\Psi(U))$. D'un point de vue théorique, la convergence de la procédure VaR-CVaR peut être obtenue à partir de la convergence faible suivante:

$$F_n^\Psi := \frac{1}{n}\sum_{k=1}^{n}\delta_{\Psi(u_k)} \overset{(\mathbb{R}^d)}{\Rightarrow} F.$$

Nous notons $D_n^*(u,\Psi)$ pour la discrépance des n premiers termes de u associé au système et définie par

$$D_n^*(u,\Psi) := \sup_{\xi\in\mathbb{R}}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\Psi(u_k)\leq\xi} - F(\xi)\right|. \tag{1.12}$$

Remarquons que dès que u est uniformément distribuée sur $[0,1]^q$ alors $D_n^*(u,\Psi)\to 0$, $n\to+\infty$.

Proposition 1.1.10. *On suppose que l'hypothèse 1.1.9 est vérifiée et que la fonction de distribution de $L=\varphi(\Psi(U))$ est lipschitz. Alors pour tout $\xi\in\mathbb{R}$*

$$(\varphi o\Psi)^{-1}\left(\,]-\infty,\xi\,]\right)\in\mathcal{M}_b,$$

où pour tout $\epsilon>0$, $b(\epsilon)=C\epsilon$ pour une certaine constante $C>0$. Si u est une suite uniformément distribuée à discrépance faible alors

$$l_n := \max_{1\leq k\leq n} kD_k^*(u,\Psi) = O\left(n^{1-\frac{1}{q}}\log(n)\right),\quad n\geq 1.$$

La vitesse de convergence obtenue ci-dessus pour la discrépance étoile correspond à l'hypothèse de moyennisation introduite dans [60]. Soit u une suite uniformément distribuée dans $[0,1]^q$, l'algorithme VaR-CVaR en utilisant la suite u est naturellement définie pour $n \geq 1$ par

$$\xi_n = \xi_{n-1} - \gamma_n K_1(\xi_{n-1}, u_n) \tag{1.13}$$
$$C_n = C_{n-1} - \gamma_n K_2(\xi_{n-1}, C_{n-1}, u_n), \tag{1.14}$$

où pour tout $\xi \in \mathbb{R}$, $u \in [0,1]^q$, $K_1(\xi, u) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(\Psi(u)) \geq \xi\}}$ et $K_2(\xi, c, u) := c - \xi - \frac{1}{1-\alpha}\left(\varphi(\Psi(u)) - \xi\right)_+$.

Theorem 1.1.11. *On suppose que l'hypothèse 1.1.9 est satisfaite et que la fonction de distribution de $L = \varphi(\Psi(U))$ est lipschitz. Soit u une suite uniformément distribuée sur $[0,1]^q$ à discrépance faible. Soit $\gamma = (\gamma_n)_{n\geq 1}$ une suite de pas réels positifs vérifiant*

$$\sum_{n\geq 1}\gamma_n = +\infty, \quad \gamma_n n^{1-\frac{1}{q}\log(n)} \longrightarrow 0, \quad \sum_{n\geq 1}\max\left(\left|\Delta\gamma_{n+1}\right|, \gamma_n^2\right) n^{1-\frac{1}{q}}\log(n) < +\infty. \tag{1.15}$$

Alors les procédures (1.13) *et* (1.14) *convergent vers ξ_α^* et C_α^*.*

Remarquons que la suite de pas $\gamma_n = \frac{c}{n}$, $c > 0$, $n \geq 1$ vérifie toujours (1.15). Des tests numériques complètent cette étude. Nous comparons notamment la vitesse de convergence entre l'algorithme VaR-CVaR avec suite à discrépance faible et avec une suite pseudo-aléatoire comme étudiée dans [6]. Les résultats montrent que lorsque la dimension q est faible (disons $q \leq 10$), l'algorithme quasi-stochastique VaR-CVaR converge plus rapidement que son homologue stochastique.

1.2 Couverture en CVaR par algorithme stochastique

Dans ce deuxième chapitre, nous nous intéressons à la couverture d'une perte (ou d'un actif) par minimisation du risque en CVaR dans un marché incomplet opérant à temps discret.

1.2.1 Introduction

Rappelons que dans un marché viable (*i.e.* en l'absence d'opportunité d'arbitrage), il existe une probabilité $\mathbb{P}^*$ (équivalente à la probabilité historique $\mathbb{P}$) sous laquelle le processus des prix des actifs négociables actualisé est une martingale. Le prix d'un actif contingent (option sur un actif négociable) est déterminé par la valeur courante d'un portefeuille autofinancé ayant presque sûrement pour valeur terminale la valeur de l'option à maturité. On dira que le marché est complet si la probabilité $\mathbb{P}^*$ est unique. Harrison et Kreps dans [43] ont montré que dans un marché complet tout actif contingent s'exprime comme une intégrale stochastique par rapport au processus des prix et peut ainsi être répliqué par un portefeuille autofinancé sans aucun risque.

Quand le marché est incomplet, le problème est plus délicat. La mesure martingale $\mathbb{P}^*$ n'est plus unique et il n'y a aucune raison pour que la valeur d'un actif contingent s'exprime comme une intégrale stochastique par rapport au processus des prix. Il existe une littérature abondante sur le problème de l'évaluation du prix des actifs contingents et de leur couverture en marché incomplet. Citons [33], [32] et [83] pour une approche par minimisation du risque quadratique, [26] pour une approche par sur-réplication, [34] et [67] pour une approche basée sur la mesure de probabilité d'entropie minimale.

Une autre méthode largement étudiée dans la littérature est basée sur la maximisation de l'utilité espérée. Cela consiste à déterminer la proportion optimale de richesse que l'investisseur doit détenir pour chaque actif en fonction de son prix afin de maximiser l'utilité espérée finale de son portefeuille. L'evaluation du prix d'un actif se fait en utilisant le principe d'indifférence d'utilité: l'utilité espérée d'un portefeuille contenant l'actif doit être égale à l'utilité de ce même portefeuille sans cet actif. Plusieurs approches et extensions ont été proposées à ce problème de maximisation. Nous nous référons à [27], [44], [47] et [65] pour plus de détails.

Dans ce chapitre, nous nous intéressons à la minimisation du risque en CVaR d'une perte (ou d'un actif) comprenant une source de risque *observable* mais *non négociable* sur les marchés de l'énergie (ou financiers). Evaluer et couvrir un actif en utilisant les mesures de risque est une approche récente qui a été étudiée par de nombreux auteurs. P. Barrieu et N. El Karoui dans [9] se sont intéressées au problème de minimisation du risque non-couvrable (à l'aide des actifs présents sur le marché financier) en utilisant les mesures de risque convexes. Les stratégies maximisant la probabilité de réussite d'une couverture sont étudiées dans [30] comme alternative aux stratégies de sur-couverture.

Dans le cadre de l'optimisation de portefeuille pour réduire la CVaR, R.T. Rockafellar et S. Uryasev dans [77] proposent une méthode afin d'estimer à un instant donné la proportion optimale que l'investisseur doit détenir en chaque actif afin de minimiser la CVaR de son portefeuille. La méthode est basée sur la programmation linéaire et consiste dans un premier temps à générer des scénarios de perte puis à les introduire comme des contraintes dans le problème de programmation linéaire. Le principal désavantage de la méthode est que le nombre de contrainte du problème est égale au nombre de scénarios simulés. Par conséquent, nous sommes très vite limités dans la pratique.

Dans ce travail, nous considérons un marché (de l'énergie ou financier) opérant à temps discret $t_0 = 0 < t_1 < \cdots < t_M = T$ et constitué d'un actif sans risque dont le prix est supposé constant et égal à 1 et de d actifs risqués négociables dont les prix à l'instant t_ℓ seront notés $X_\ell = (X_\ell^i)_{1\leq i\leq d}$. La perte du portefeuille subie à l'instant T (ou le payoff du produit dérivé à l'instant T) qu'on désire couvrir est une variable aléatoire réelle L définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. La source d'incomplétude du marché vient de la présence dans L d'un processus $Z := (Z_\ell)_{1\leq \ell\leq M}$ qui est *observable* mais *non négociable*, cela induit une source de risque qui n'est pas couvrable à l'aide des actifs présents sur le marché. Par exemple, sur le marché de l'électricité ou du gaz, la perte pour un fournisseur d'énergie peut être due à une consommation de ses clients (entreprises ou particuliers) en électricité ou en gaz. Généralement, cette consommation dépend de la température qui est un processus observable mais non négociable sur les marchés. De plus, on peut consid-

érer que les prix de l'électricité dépendent eux aussi de la température de sorte que le processus Z peut intervenir dans la dynamique des prix. Ce type de dépendance vis à vis d'un processus observable mais non négociable est particulièrement pertinente sur les marchés financiers et de l'énergie (par exemple, volatilité stochastique et temps de défaut sur les marchés financiers; température et autres aléas climatiques sur les marchés de l'énergie). Nous notons $\mathbb{G} = (\mathcal{G}_\ell)_{0\leq\ell\leq M}$ la filtration engendrée par les processus X et Z, *i.e.* $\mathcal{G}_\ell = \sigma\{X_i, Z_i; 0 \leq i \leq \ell\}$.

Afin de réduire son risque (ou couvrir son actif contingent), nous considérons que l'investisseur utilise une stratégie dynamique autofinancée. A temps discret, celle-ci peut être représentée par un processus $\theta = (\theta_\ell)_{1\leq\ell<M}$ adapté à la filtration $\mathbb{G}$, tel que $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{G}_\ell)$, où $L^0_{\mathbb{R}^d}(\mathcal{G})$ est l'espace des variables aléatoires $\mathcal{G}$-mesurable et finies $p.s.$ à valeurs dans $\mathbb{R}^d$. Dans une telle stratégie, θ_ℓ peut être vu comme la quantité d'actifs investie à l'instant t_ℓ. Le gain à l'instant T d'une telle stratégie partant d'une richesse initiale nulle est donné par l'intégrale stochastique discrète $\sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell$ où $\Delta X_\ell := X_\ell - X_{\ell-1}$.

Le problème pour l'investisseur est de trouver une stratégie autofinancée θ^*_α qui minimise le risque résiduel de la perte du portefeuille. Nous nous intéressons donc au problème de minimisation de la CVaR:

$$\inf_{\theta\in\mathcal{A}} \mathrm{CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right), \tag{1.16}$$

où

$$\mathcal{A} = \left\{\theta \mid \theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{G}_\ell),\ \ell = 0, \cdots,\ M-1\right\}$$

est l'ensemble des stratégies admissibles sous l'hypothèse crucial que le processus de prix $(X_\ell)_{0\leq\ell\leq M}$ est une $(\mathcal{G},\mathbb{P})$-martingale.

Afin de mesurer le risque à un instant t_k, nous introduisons une nouvelle mesure de risque que nous noterons $\mathcal{G}_k$-CVaR qui est une version dynamique et donc aléatoire de la CVaR et qui est basée sur la représentation introduite par Rockafellar et Uryasev (voir Proposition 1.1.1). Nous considérons une sous-tribu $\mathcal{F} \subset \mathcal{A}$.

Definition 1.2.1. Soit L tel que $\mathbb{E}[L_+] < +\infty$. La $\mathcal{F}$-CVaR est la mesure de risque aléatoire définie par

$$\mathcal{F}\text{-CVaR}_\alpha(L) := \operatorname*{ess\,inf}_{\xi\in L^0_{\mathbb{R}}(\mathcal{F})} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\xi)_+ \,|\mathcal{F}\right].$$

Par construction, il est clair qu'elle satisfait les propriétés suivantes

1. Sous-additivité: Pour tout $L,\ L' \in L^1(\mathbb{P})$, $\mathcal{F}\text{-CVaR}_\alpha(L+L') \leq \mathcal{F}\text{-CVaR}_\alpha(L) + \mathcal{F}\text{-CVaR}_\alpha(L')$.

2. Homogénéité positive: Soit $\lambda \in L^0_{\mathbb{R}}(\mathcal{F})$ tel que $\lambda \geq 0$ $p.s.$, $\mathcal{F}\text{-CVaR}_\alpha(\lambda L) = \lambda \times \mathcal{F}\text{-CVaR}_\alpha(L)$.

3. Invariance par translation: Pour tout $Z \in L^0_{\mathbb{R}}(\mathcal{F})$, $\mathcal{F}\text{-CVaR}_\alpha(L+Z) = Z + \mathcal{F}\text{-CVaR}_\alpha(L)$.

4. Monotonicité: Pour tout $L,\ L' \in L^1(\mathbb{P})$ tel que $L \leq L',\ \ \mathcal{F}\text{-CVaR}_\alpha(L) \leq \mathcal{F}\text{-CVaR}_\alpha(L')$.

Ainsi, afin d'estimer cette quantité qui apparaît comme une espérance conditionnelle, nous utilisons une formule de quadrature basée sur la quantification optimale.

Pour plusieurs raisons (coûts de transaction, coûts d'entrée et de sortie liés au stockage du gaz, impossibilité de stocker de l'électricité, horizon de temps court, ...), un investisseur peut ne pas vouloir intervenir sur les marchés chaque jour. Il peut être intéressé par une couverture approximative de son profil de risque. Pour ces raisons, nous nous intéressons d'abord aux stratégies de couverture à un pas. Une telle stratégie décidée à l'instant t_{ℓ_0}, $\ell_0 = 0, \cdots, M-1$, est obtenue en fixant définitivement la quantité initiale d'actifs contenue dans le portefeuille $\theta_\ell \equiv \theta_{\ell_0}$, pour $\ell = \ell_0, \cdots, M-1$. Elle est donc représentée par une seule variable aléatoire $\theta_{\ell_0} \in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0})$. Le risque de l'investisseur à l'instant t_{ℓ_0} est mesuré par la quantité aléatoire connue uniquement à l'instant t_{ℓ_0}: $\mathcal{G}_{\ell_0}$-$\mathrm{CVaR}_\alpha\left(L - \theta_{\ell_0}.\left(X_M - X_{\ell_0}\right)\right)$. A la date 0, l'investisseur l'estime en calculant $\mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-}\mathrm{CVaR}_\alpha\left(L - \theta_{\ell_0}.\left(X_M - X_{\ell_0}\right)\right)\right]$. Cette quantité représente un risque forward, *i.e.* c'est la meilleur approximation à la date 0 du risque à la date t_{ℓ_0} alors que la quantité $\mathrm{CVaR}_\alpha\left(L - \theta_{\ell_0}.\left(X_M - X_{\ell_0}\right)\right)$ représente le risque à la date 0. Par conséquent, il y a deux problèmes d'optimisation possibles pour l'investisseur.

Le premier consiste à minimiser le risque forward, *i.e.* l'espérance du profile de risque à la date t_{ℓ_0} de la perte du portefeuille en utilisant une stratégie auto-financée à un pas problème et une richesse initiale nulle

$$\inf_{\theta_{\ell_0} \in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0}, \mathbb{P})} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-}\mathrm{CVaR}_\alpha\left(L - \theta_{\ell_0}.\left(X_M - X_{\ell_0}\right)\right)\right]. \tag{1.17}$$

Le second consiste à minimiser le risque à la date 0 des pertes du portefeuille en utilisant un critère de CVaR statique et le même type de stratégie

$$\inf_{\theta_{\ell_0} \in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0}, \mathbb{P})} \mathrm{CVaR}_\alpha\left(L - \theta_{\ell_0}.\left(X_M - X_{\ell_0}\right)\right). \tag{1.18}$$

La VaR_α et la CVaR_α sont des mesures de risque disymétriques contrairement à l'écart-type. En utilisant la couverture en CVaR, nous cherchons à modifier la distribution de la perte L de manière disymétrique, *i.e.* nous désirons réduire davantage la queue de distribution à droite (celle qui correspond aux pertes extrêmes) que celle à gauche qui correspond aux pertes faibles voire aux gains potentiels. C'est la différence fondamentale entre la couverture en CVaR et la couverture par minimisation du risque quadratique.

Nous étudions l'existence de solutions aux problèmes (1.16), (1.17) et (1.18).

1.2.2 Aspects théoriques de la couverture en CVaR

Dans un premier temps, nous nous intéressons aux propriétés de la $\mathcal{G}_k$-CVaR_α afin de mieux comprendre comment le risque du portefeuille détenu par l'investisseur évolue jusqu'à la maturité T.

Proposition 1.2.1. *Nous posons $M = +\infty$ pour ce résultat. Soit Y une variable aléatoire réel tel que $\mathbb{E}[Y_+] < +\infty$. La séquence $\left(\mathcal{G}_k\text{-}\mathrm{CVaR}_\alpha(Y)\right)_{k\geq 0}$ est une surmartingale. De plus, si Y est $\mathcal{G}_\infty$-mesurable, où $\mathcal{G}_\infty = \vee_k \mathcal{G}_\infty$ alors on a*

$$\mathcal{G}_n\text{-}\mathrm{CVaR}_\alpha(Y) \xrightarrow{p.s.} Y,\ n \to +\infty.$$

Cela a pour conséquence immédiate que la séquence $\left(\mathbb{E}\left[\mathcal{G}_k\text{-CVaR}_\alpha(Y)\right]\right)_{0\leq k\leq M}$ est décroissante et donc l'estimation à la date 0 du risque dynamique de l'investisseur décroit au cours du temps. Nous dirons que la mesure de risque aléatoire $\mathcal{G}_k\text{-CVaR}_\alpha(Y)$ est cohérente.

Nous continuons ensuite avec un corollaire concernant le portefeuille avec stratégie de couverture dynamique.

Corollary 1.2.2. *Soit L tel que $\mathbb{E}\left[L_+\right] < +\infty$. On suppose qu'il existe $p' > 1$ tel que $\Delta X_\ell \in L^{p'}(\mathbb{P})$ pour $\ell = 1, \cdots, M$. Soit $\theta \in \mathcal{A}$ tel que $\theta_\ell \in L^p(\mathbb{P})$, pour $\ell = 0, \cdots, M-1$ avec $p = \frac{p'}{p'-1}$. Alors la séquence*

$$\left(\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right)_{0\leq k\leq M} \quad \textit{est une surmartingale}$$

et vérifie pour $k \in \{0, \cdots, M-1\}$,

$$\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) = \mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) - \sum_{\ell=1}^{k} \theta_{\ell-1}\Delta X_\ell \tag{1.19}$$

Si X est une $(\mathcal{G}, \mathbb{P})$-martingale, il en découle l'identité suivante

$$\mathbb{E}\left[\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right] = \mathbb{E}\left[\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right],$$

qui signifie que l'estimation à la date 0 du risque à la date t_k ne dépend pas des décisions prises avant cette dernière date.

On considère une sous-tribu $\mathcal{F} \subseteq \mathcal{A}$ avec la possibilité par la suite (lorsque nous proposerons des procédures numériques d'estimation de stratégie optimale) de poser $\mathcal{F} = \mathcal{G}_{\ell_0}$, $\ell_0 = 0, \cdots, M-1$. On pose $X = X_M - X_{\ell_0}$ et nous nous intéressons aux deux problèmes

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right], \tag{1.20}$$

et

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \text{CVaR}_\alpha\left(L - \theta.X\right). \tag{1.21}$$

Remarquons que le problème (1.21) peut être écrit

$$\inf_{\xi \in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right], \tag{1.22}$$

ainsi par conséquent, nous résolvons dans un premier temps le problème suivant

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right]. \tag{1.23}$$

Sans perte de généralité, nous pouvons supposer que $\xi = 0$ et $\alpha = 0$. Ainsi le problème (1.23) est équivalent à minimiser le risque "short fall"

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\left(L - \theta.X\right)_+\right]. \tag{1.24}$$

Soit $\Pi(\mathrm{d}y, \mathrm{d}x) = \Pi(\omega, \mathrm{d}y, \mathrm{d}x)$ une version régulière de la loi conditionnelle de (L, X) sachant $\mathcal{F}$.

Afin d'établir l'existence de solutions aux deux problèmes ci-dessus, nous faisons l'hypothèses suivantes concernant cette loi conditionnelle

Hypothèse 1.2.3. *Cas du risque statique*

i) La distribution de L et de X satisfait $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $X \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$.

ii) $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathbb{E}\left[(u.X)_+ \mid \mathcal{F}\right] > 0$ *p.s.*

Hypothèse 1.2.4. *Cas du risque forward*

i) La distribution de L et de X satisfait $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $X \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$.

ii) $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathcal{F}\text{-CVaR}_\alpha\left(u.X\right) > 0$ *a.s.*

Proposition 1.2.5. *Soit V_f et V_s deux fonctions définies sur $\Omega \times \mathbb{R} \times \mathbb{R}^d$ et $\Omega \times \mathbb{R}^d$ par*

$$V_f(\omega, \xi, \theta) = \int v_f(\xi, \theta, y, x)\Pi(\omega, \mathrm{d}x, \mathrm{d}y), \tag{1.25}$$

$$V_s(\omega, \xi, \theta) = \int v_s(\theta, y, x)\Pi(\omega, \mathrm{d}x, \mathrm{d}y) \tag{1.26}$$

où

$$v_f(\xi, \theta, y, x) = \xi + \frac{1}{1-\alpha}\left(y - \theta.x - \xi\right)_+, \tag{1.27}$$

et

$$v_s(\theta, y, x) = \left(y - \theta.x\right)_+, \tag{1.28}$$

Alors, on a

i) Risque statique: *Supposons que l'hypothèse 1.2.3 est vérifiée. Alors pour tout $\omega \in \Omega$, la fonction $V_s(\omega, .)$ est convexe, Lipschitz continue et $\lim_{|\theta|\to+\infty} V_s(\omega, \theta) = +\infty$. De plus, on a*

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\left(L - \theta.X\right)_+\right] = \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\left(L - \theta.X\right)_+ \middle| \mathcal{F}\right]\right], \tag{1.29}$$

et

$$\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\left(L - \theta.X\right)_+ \middle| \mathcal{F}\right](\omega) = \min_{\theta \in \mathbb{R}^d} V_s(\omega, \theta). \tag{1.30}$$

ii) Risque forward: *Supposons que l'hypothèse 1.2.4 est vérifiée. Alors, pour tout $\omega \in \Omega$, la fonction $V_f(\omega, ., .)$ est convexe, continue et $\forall \xi \in \mathbb{R}$, $\lim_{|(\xi,\theta)|\to+\infty} V_f(\omega, \xi, \theta) = +\infty$. De plus on a,*

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right] = \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ \xi \in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ \middle| \mathcal{F}\right]\right] \tag{1.31}$$

et

$$\operatorname*{ess\,inf}_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ \xi\in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\theta.X-\xi\right)_+\middle|\mathcal{F}\right](\omega)=\min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\omega,\xi,\theta). \tag{1.32}$$

Le membre de droite des deux équations (1.30) et (1.32) montre que les deux problèmes d'optimisation (1.20) et (1.21) s'écrivent

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\theta.X\right)\right]=\mathbb{E}\left[\min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V(\xi,\theta)\right], \tag{1.33}$$

et,

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F})} \text{CVaR}_\alpha\left(L-\theta.X\right)=\inf_{\xi\in\mathbb{R}} \mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V(\xi,\theta)\right]. \tag{1.34}$$

Par conséquent, pour tout $\omega\in\Omega$, nous devons résoudre des problèmes d'optimisation déterministes. Le résultat suivant permet de caractériser ces minimas et nous permettra par la suite d'en déduire des procédures numériques afin d'estimer les stratégies optimales.

Proposition 1.2.6. *Supposons que l'hypothèse 1.2.3 est vérifiée. Alors, pour tout* $\xi\in\mathbb{R}$

$$\operatorname{Arg\,min} V_f(\xi,.)=\left\{\theta\in\mathbb{R}^d|\ \nabla_\theta V_f(\xi,\theta)=0\right\}\neq\varnothing.$$

Supposons que l'hypothèse 1.2.4. est vérifiée alors

$$\operatorname{Arg\,min} V_f=\left\{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d|\ \nabla_{(\xi,\theta)} V_f(\xi,\theta)=0\right\}\neq\varnothing.$$

où le gradient de V_f *peut être représenté pour tout* $(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d$ *par*

$$\nabla_{(\xi,\theta)} V_f(\xi,\theta)=\int \nabla_{(\xi,\theta)} v_f(\xi,\theta,y,x)\Pi(\mathrm{d}x,\mathrm{d}y) \tag{1.35}$$

et,

$$\nabla_\theta V_f(\xi,\theta)=\int \nabla_\theta v_f(\xi,\theta,y,x)\Pi(\mathrm{d}x,\mathrm{d}y). \tag{1.36}$$

De plus, $\xi\mapsto\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right]$ *est Lipschitz continue, convexe, et* $\lim_{|\xi|\to+\infty}\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right]=+\infty$. *Par conséquent,* (1.20) *et* (1.21) *admettent des solutions.*

Concernant les stratégies de couverture dynamique, on considère de la même façon le problème plus général suivant

$$\inf_{\theta\in\mathcal{A}_\mathcal{F}} \text{CVaR}_\alpha\left(L-\sum_{\ell=1}^M \theta_{\ell-1}.\Delta X_\ell\right)=\inf_{\xi\in\mathbb{R}}\inf_{\theta\in\mathcal{A}_\mathcal{F}} \mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\sum_{\ell=1}^M \theta_{\ell-1}.\Delta X_\ell-\xi\right)_+\right] \tag{1.37}$$

où $\mathcal{A}_\mathcal{F}$ est l'ensemble des séquences $\theta=(\theta_0,\cdots,\theta_{M-1})$ tel que $\theta_\ell\in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\cdots\mathcal{F}_{M-1}\subseteq\mathcal{A}$ sont des sous-tribus avec la possibilité de fixer $\mathcal{F}_\ell$ to $\mathcal{G}_\ell$.

Afin d'obtenir des résultats similaires au cadre "un pas", nous considérons une famille de lois conditionnelles $(\Pi_\ell)_{0\leq\ell\leq M-1}$ où $\Pi_\ell(\mathrm{d}y,\mathrm{d}x)=\Pi_\ell(\omega,\mathrm{d}y,\mathrm{d}x)$ est une version régulière de la loi conditionnelle du vecteur $(L,\Delta X_1,\cdots,\Delta X_M)$ sachant $\mathcal{F}_\ell$. Nous nous plaçons sous l'hypothèse suivante:

Hypothèse 1.2.7. *i) La distribution de* $(L, \Delta X_1, \cdots, \Delta X_M)$ *satisfait* $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $\Delta X_\ell \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$, $\ell = 1, \cdots, M$

ii) $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathbb{E}\left[(u.\Delta X_\ell)_+ \mid \mathcal{F}_{\ell-1}\right] > 0$ *p.s.*

Avant d'établir l'existence d'une solution au problème (1.37), nous le décomposons en deux sous-problèmes. Dans une premier temps, nous intéressons aux problèmes de contrôle stochastiques suivant

$$\inf_{\theta \in \mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+\right], \quad \text{pour tout} \quad \xi \in \mathbb{R}. \tag{1.38}$$

En utilisant des idées similaires à celles utilisées dans le principe de la programmation dynamique (fonctions de Bellman, etc), nous montrons l'existence d'une stratégie optimale. Dans cet esprit, nous construisons la solution de (1.38) de manière "backward", pas à pas de la date ℓ à la date $\ell - 1$. Sans perte de généralité, on pose $\xi = 0$ et $\alpha = 0$. On remarque que le problème (1.38) peut s'écrire

$$\inf_{\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell,\mathbb{P}),\ell=0,\cdots,M-2} \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta_{M-1} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1},\mathbb{P})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)_+ \middle| \mathcal{F}_{M-1}\right]\right]. \tag{1.39}$$

Donc, nous pouvons commencer par résoudre le problème

$$\operatorname*{ess\,inf}_{\theta_{M-1} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)_+ \middle| \mathcal{F}_{M-1}\right](\omega) = \min_{\theta_{M-1} \in \mathbb{R}^d} V_{M-1}(\omega, \theta) \ \ p.s. \tag{1.40}$$

où V_{M-1} est définie pour tout $\omega \in \Omega$, $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\ell = 1, \cdots, M-21$, par

$$\begin{aligned} V_{M-1}(\omega, \theta_{0:M-2}, \theta_{M-1}) &:= \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M-1} \theta_{\ell-1}.\Delta X_\ell - \theta_{M-1}.\Delta X_M\right)_+ \middle|\ \mathcal{F}_{M-1}\right](\omega) \\ &= \int w(\theta_{0:M-2}, \theta_{M-1}, y, x)\Pi_{M-1}(\omega, \mathrm{d}x, \mathrm{d}y) \end{aligned} \tag{1.41}$$

avec

$$w(\theta, y, x) = \left(y - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta x_\ell\right)_+. \tag{1.42}$$

En utilisant des arguments similaires à ceux utilisés dans le cadre statique, nous montrons que pour tout $\omega \in \Omega$, $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\ell = 0, \cdots, M-1$, pour tout $\xi \in \mathbb{R}$, la fonction $\theta_{M-1} \mapsto V_{M-1}(\omega, \theta_{0:M-2}, \theta_{M-1})$ est convexe, Lipschitz et satisfait $\lim_{|\theta_{M-1}| \to +\infty} V_{M-1}(\omega, \theta_{0:M-2}, \theta_{M-1}) = +\infty$. Par conséquent, le problème (1.40) admet une solution que l'on notera $\tilde{\theta}_{M-1,\alpha}$. Ensuite si $\tilde{\theta}_{\alpha,\ell:M-1} := (\tilde{\theta}_{\alpha,\ell}, \cdots, \tilde{\theta}_{\alpha,M-1})$ est la solution construite jusqu'à l'étape ℓ, on montre qu'il possible de construire $\tilde{\theta}_{\alpha,\ell-1}$ au pas $\ell - 1$, en résolvant le problème d'optimisation suivant

$$\operatorname*{ess\,inf}_{\theta_{\ell-1} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{\ell-1},\mathbb{P})} \mathbb{E}\left[V_\ell(\theta_{0:\ell-2}, \theta_{\ell-1}) \middle| \mathcal{F}_{\ell-1}\right](\omega) = \min_{\theta_{\ell-1} \in \mathbb{R}^d} V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1}) \ \ a.s. \tag{1.43}$$

où pour tout $\theta_{\ell-1} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{\ell-1}, \mathbb{P})$, les fonctions V_ℓ and $V_{\ell-1}$ sont définies par

$$V_\ell(\omega, \theta_{0:\ell-2}, \theta_{\ell-1}) := \mathbb{E}\left[\left(L - \sum_{k=1}^{\ell-1} \theta_{k-1}.\Delta X_k - \theta_{\ell-1}.\Delta X_\ell - \sum_{k=\ell+1}^{M} \tilde{\theta}_{k-1}.\Delta X_k\right)_+ \middle| \mathcal{F}_\ell\right](\omega)$$
$$= \int w(\theta_{0:\ell-2}, \theta_{\ell-1}, \tilde{\theta}_{\ell:M-1}, y, x)\Pi_\ell(\omega, \mathrm{d}x, \mathrm{d}y) \tag{1.44}$$

et,

$$V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1}) = \int w(\theta_{0:\ell-2}, \theta_{\ell-1}, \tilde{\theta}_{\ell:M-1}, y, x)\Pi_{\ell-1}(\omega, \mathrm{d}x, \mathrm{d}y). \tag{1.45}$$

Une fois que la stratégie optimale $\tilde{\theta}$ est construite donc également la solution θ^*_α de (1.38), on revient au problème global (1.37) qui s'écrit

$$\inf_{\xi\in\mathbb{R}} \inf_{\theta\in\mathcal{A}_\mathcal{F}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+\right]$$

dont on établit l'existence d'un "ξ" optimal grâce à une uniforme Lipschitzité en la variable "θ". La proposition suivante implique que le problème (1.37) admet une solution optimale $(\xi^*_\alpha, \theta^*_\alpha) \in \mathbb{R} \times \mathcal{A}_\mathcal{F}$.

Proposition 1.2.8. *Supposons que l'hypothèse 1.2.7 est vérifiée. Alors,* (1.39) *est satisfaite, le problème* (1.40) *admet une solution et pour $\ell = M-1, \cdots, 1$, il existe $\theta^*_{\alpha,\ell-1}$ $\mathcal{F}_{\ell-1}$-mesurable, solution du problème* (1.43).
De plus, la fonction $\xi \mapsto \inf_{\theta\in\mathcal{A}_\mathcal{F}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+\right]$ est convexe, Lipschitz et satisfait $\lim_{|\xi|\to+\infty} \inf_{\theta\in\mathcal{A}_\mathcal{F}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+\right] = +\infty$. Ainsi le problème (1.37) *admet une solution.*

1.2.3 Aspects numériques de la couverture en CVaR

Nous nous intéressons dans une seconde partie aux aspects numériques de la couverture en CVaR dans un cadre markovien. Nous étudions des méthodes numériques afin d'estimer les stratégies optimales solutions de (1.17), (1.18) et (1.16) sous l'hypothèse

Hypothèse 1.2.9. *(Cadre Markovien)*

1. *Le processus $(X_\ell, Z_\ell)_{0\le\ell\le M}$ à valeurs dans $\mathbb{R}^d \times \mathbb{R}^q$ est markovien par rapport à la filtration $\mathbb{G}$.*

2. *Le processus $(X_\ell)_{0\le\ell\le M}$ est une $(\mathbb{G}, \mathbb{P})$-martingale.*

Sur les marchés de l'énergie, par exemple ceux du gaz et de l'électricité, le prix des actifs X présents sont essentiellement les contrats *day-ahead* et les contrats *forward*. Du fait de l'impossibilité de stocker de l'électricité et du coût très élevé pour le stockage du gaz, il est impossible d'établir une stratégie de gestion autofinancée sur

les contrats day-ahead et seules les contrats forwards sont disponibles à cette fin. Le prix de ces derniers est souvent modélisé par une martingale sous la probabilité historique à l'aide de modèles de type Heath-Jarrow-Morton.

Dans un premier temps, nous développons un algorithme stochastique pour les deux problèmes à un pas. Soit $\ell_0 \in 0, \cdots, M-1$. D'un point de vue modélisation, nous supposons que la perte s'écrit $L = \phi(X, Z)$ et qu'il existe deux fonction $F : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}$ et $G : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}^d$ tel que

$$X_M - X_{\ell_0} = G(X_{\ell_0}, Z_{\ell_0}, U_{\ell_0} + 1) \text{ and } L = F(X_{\ell_0}, Z_{\ell_0}, U_{\ell_0+1})$$

où U_{ℓ_0+1} est un vecteur aléatoire de dimension r_{ℓ_0} indépendant de $\mathcal{G}_{\ell_0} := \sigma(X_{\ell_0}, Z_{\ell_0})$. On note U pour U_{ℓ_0+1}. Dans ce cadre markovien, la fonction (1.25) peut s'écrire pour tout $(x, z) \in \mathbb{R}^d \times \mathbb{R}^q$

$$V_f(\xi, \theta, x, z) = \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(F(x, z, U) - \theta.G(x, z, U) - \xi)_+\right],$$

donc (1.32) devient

$$\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0}), \xi \in L^0_{\mathbb{R}}(\mathcal{G}_{\ell_0})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}(L - \theta.X - \xi)_+ \middle| \mathcal{G}_{\ell_0}\right] = \left(\min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi, \theta, x, z)\right)_{(x,z)=(X_{\ell_0}, Z_{\ell_0})} \tag{1.46}$$

Par conséquent, afin de résoudre le problème global (1.17) nous devons résoudre le problème local qui apparaît dans le membre de droite de l'égalité ci-dessus pour chaque $(X_{\ell_0}(\omega), Z_{\ell_0}(\omega))$. Ensuite, nous devons estimer la quantité suivante

$$\mathbb{E}\left[\left(\inf_{(\theta,\xi)\in\mathbb{R}^d\times\mathbb{R}} V_f(\xi, \theta, x, z)\right)_{|x=X_{\ell_0}, z=Z_{\ell_0}}\right]. \tag{1.47}$$

Pour cela, nous proposons d'utiliser une grille de quantification optimale du processus (X_{ℓ_0}, Z_{ℓ_0}), $\Gamma_{\ell_0} := \Gamma^{N_{\ell_0}}_{\{X_{\ell_0}, Z_{\ell_0}\}} = \left((x^1_{\ell_0}, z^1_{\ell_0}), \cdots, (x^{N_{\ell_0}}_{\ell_0}, z^{N_{\ell_0}}_{\ell_0})\right)$ de taille N_{ℓ_0}.

Pour chaque noeud $(x^j_{\ell_0}, z^j_{\ell_0})$ de la grille, ce qui correspond à un état probable de la variable aléatoire (X_{ℓ_0}, Z_{ℓ_0}), nous estimons la stratégie $\theta^*_{\ell_0}(x^j_{\ell_0}, z^j_{\ell_0})$ ainsi que la VaR_α locale $\xi^*_\alpha(x^j_{\ell_0}, z^j_{\ell_0})$ et la CVaR_α locale $CV^*_\alpha(x^j_{\ell_0}, z^j_{\ell_0})$ à l'aide d'un algorithme stochastique $(\xi_n, \theta_n, C_n)_{n\geq 1}$. Puis, nous estimons la CVaR globale du portefeuille C^*_α, *i.e.* la quantité (1.47) à l'aide de la formule de quadrature suivante

$$\sum_{j=1}^{N_k} CV^*_\alpha(x^j_k, z^j_k)\mathbb{P}\left((X_k, Z_k) \in C_j(x_k, z_k)\right), \tag{1.48}$$

où la suite $(C_j(x_{\ell_0}, z_{\ell_0}))_{1\leq j\leq N_{\ell_0}}$ est la partition de Voronoï associée à la grille (x_{ℓ_0}, z_{ℓ_0}).

Comme pour l'algorithme d'estimation du couple VaR-CVaR étudiée dans [6], afin d'obtenir la meilleure vitesse de convergence asymptotique, nous utilisons le principe de moyennisation de Ruppert & Polyak. Nous nous intéressons ensuite aux stratégies autofinancées dynamiques et nous en proposons quatre différentes afin d'approcher la stratégie optimale solution de (1.16):

- **Crude CVaR hedging algorithm (C.H.)**: C'est la procédure qui est la plus naturelle. Elle consiste à résoudre directement le problème en (1.16) en généralisant l'algorithme développé dans le cadre statique. Pour être plus précis, nous quantifions le processus (X, Z) à chaque date t_k, pour $1 \leq k \leq M$. Le nombre d'actifs à détenir pendant la période $(k, k+1]$, θ_k^* dépend de toute la trajectoire du processus (X, Z). Nous faisons l'approximation consistant à faire dépendre θ_k^* uniquement de (X_k, Z_k). À l'étape n de l'algorithme stochastique, la stratégie dynamique optimale est estimée par la variable aléatoire $\theta_n = (\theta_{0,n}, \cdots, \theta_{M-1,n})$ et pour $k = 1, \cdots, M$. La VaR et la CVaR sont également calculées par algorithme stochastique mais, contrairement au cas statique, elles sont estimées de manière globable et non plus locale, *i.e.* qu'elles ne dépendent plus du noeud de l'arbre de quantification considéré.

  Le principale désavantage de cette méthode réside dans la dimension de l'algorithme stochastique mis en oeuvre. Elle est égale à $D := 2 + d + \sum_{k=2}^{M} d \times N_k$. Lorsque celle-ci est faible ($D \leq 100$) on observe un très bon comportement de l'algorithme. Mais dès que la dimension devient grande $D \geq 150, 200$, ce qui est souvent le cas en pratique, l'algorithme ne converge plus et l'efficacité de la méthode diminue grandement. Les trois autres méthodes que nous proposons palient ce problème tout en approchant la solution du problème (1.16).

- **Backward Dynamic hedging algorithm (B.H.)**: Cette stratégie est basée sur (1.19) et consiste en une résolution backward. Si l'on considère M dates sur la période, alors afin de se couvrir au mieux à la dernière date, *i.e.* à la date t_{M-1}, l'équation (1.19) nous dit qu'il faut résoudre le problème d'optimisation suivant

$$\inf_{\theta \in \mathcal{A}} \mathbb{E}\left[\mathcal{G}_{M-1}\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right] = \inf_{\theta_{M-1} \in L^0_{\mathbb{R}^d}(\mathcal{G}_{t_{M-1}})} \mathbb{E}\left[\mathcal{G}_{M-1}\text{-CVaR}_\alpha\left(L - \theta_{M-1}.\Delta X_M\right)\right].$$

  Remarquons que c'est un problème d'optimisation statique à la date t_{M-1}. La stratégie optimale θ_{M-1}^b peut être facilement estimée à l'aide d'un algorithme stochastique. Ensuite, nous pouvons reculer d'un pas et résoudre le problème à la date t_{M-2} connaissant la solution optimale à la date t_{M-1}, *i.e.*

$$\inf_{\theta_{M-2} \in L^0_{\mathbb{R}^d}(\mathcal{G}_{t_{M-2}})} \mathbb{E}\left[\mathcal{G}_{M-2}\text{-CVaR}_\alpha\left(L - \theta_{M-1}^b.\Delta X_M - \theta_{M-2}\Delta X_{M-1}\right)\right],$$

  en utilisant encore une fois l'algorithme statique. On poursuit ainsi jusqu'à obtenir la stratégie optimale backward $\theta^b = \left(\theta_k^b\right)_{1 \leq k \leq M}$.

- **Martingale Decomposition of L (M.D.H.)**: Cette stratégie est basée sur une décomposition de la perte L à maturité T comme somme d'accroissements de martingale, *i.e.*

$$L = \mathbb{E}[L] + \sum_{\ell=1}^{M} \tilde{\Delta} L_\ell, \tag{1.49}$$

où $\tilde{\Delta}L_\ell = \mathbb{E}\left[L|\,\mathcal{G}_\ell\right] - \mathbb{E}\left[L|\,\mathcal{G}_{\ell-1}\right]$, $1 \leq \ell \leq M$, définie une séquence d'accroissements de martingale par rapport à $\mathbb{G}$. Maintenant, on utilise la propriété de sous-additivité de la CVaR,

$$\inf_{\theta\in\mathcal{A}} \mathrm{CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}\Delta X_\ell\right) \leq \mathbb{E}\left[L\right] + \sum_{\ell=1}^{M} \inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1})} \mathrm{CVaR}_\alpha\left(\tilde{\Delta}L_\ell - \theta_{\ell-1}\Delta X_\ell\right). \tag{1.50}$$

Nous pouvons maintenant résoudre chaque problème locaux apparaissant dans le membre de droite de (1.50) séparément. Cependant, il n'y a aucune raison pour que la stratégie ainsi obtenue soit optimale pour le problème global (1.16). D'un point de vue numérique, nous remarquons que lorsque le nombre de dates M est élevé ($M \geq 10$ en pratique) cette stratégie fournit de bons résultats. D'encore meilleurs résultats sont obtenus en la modifiant légèrement. Pour cela nous utilisons l'inégalité

$$\inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1})} \mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-}\mathrm{CVaR}_\alpha\left(\tilde{\Delta}L_\ell - \theta_{\ell-1}\Delta X_\ell\right)\right] \leq \inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1})} \mathrm{CVaR}_\alpha\left(\tilde{\Delta}L_\ell - \theta_{\ell-1}\Delta X_\ell\right), \tag{1.51}$$

ce qui nous conduit à résoudre le problème

$$\sum_{\ell=1}^{M} \inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1})} \mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-}\mathrm{CVaR}_\alpha\left(\tilde{\Delta}L_\ell - \theta_{\ell-1}\Delta X_\ell\right)\right],$$

i.e. une somme de problèmes locaux.

- **Classical Decomposition of L (C.D.H.)**: Cette stratégie est similaire à la précédente. La seule différence réside dans la décomposition de L qui est içi classique, *i.e.*

$$L = L_0 + \sum_{\ell=1}^{M} L_\ell - L_{\ell-1}.$$

Enfin, à l'instar de l'algorithme stochastique VaR-CVaR développé dans [6], nous proposons de combiner tous ces algorithmes avec des méthodes de réduction de variance afin d'accélérer la convergence asymptotique.

La première méthode que nous proposons est basée sur un algorithme récursif d'échantillonnage préférentiel que nous combinons adaptativement avec la procédure d'estimation des quantités $(\xi^*_\alpha, \theta^*_\alpha, C^*_\alpha)$. Notons que cet algorithme est plus complexe que dans le cas de la VaR-CVaR car une nouvelle composante apparaît içi: la stratégie estimée θ_n au pas n.

La seconde méthode est basée sur la variable de contrôle $X_M - X_k$ (dans le cas de la stratégie statique) ou ΔX_ℓ (dans le cas de la stratégie dynamique), $\ell = 1, \cdots, M$.

Puisque le processus X est une martingale, l'espérance de celle-ci est nulle. Nous établissons alors une formule pour calculer les poids optimaux qui réduisent la variance asymptotique dans le TCL vérifiée par la procédure $(\xi_n, \theta_n, C_n)_{n\geq 1}$. Ces poids peuvent être estimés de manière adaptative avec les quantités $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$.

Pour terminer ce chapitre, nous étudions l'efficacité des différentes stratégies étudiées ainsi que des méthodes de réduction de variance proposées sur des portefeuilles liés au marché de l'énergie (qui est fortement incomplet) où nous prenons l'exemple de la température comme processus observable mais non négociable. Nous traçons les histogrammes des pertes avant et après couverture afin de bien saisir comment la couverture en CVaR transforme et désymétrise la distribution des pertes.

1.3 Modélisation Conjointe des prix spot du gaz et de l'électricité

La récente évolution des marchés de l'énergie en Europe ces dernières années a confronté ses différents acteurs à de nombreux problèmes de modélisation et de valorisation. Comprendre et modéliser l'évolution des prix des énergies est maintenant devenu un enjeu stratégique pour les énergéticiens.

Il existe essentiellement trois types de produit que l'on peut échanger sur ces nouvelles places boursières.

Le premier est le contrat forward (ou futures). C'est un contrat qui permet à son détenteur de recevoir dans le futur une certaine quantité d'énergie fixée à l'avance durant toute une période comprise entre deux dates T_1 et T_2 (une semaine, un mois, un trimestre, une année, ...). Le prix à l'instant $t < T_1 \leq T_2$ d'un tel contrat est noté $F(t, T_1, T_2)$. Le deuxième est le contrat spot ou day-ahead. C'est un contrat qui livre pour le lendemain une quantité d'énergie connue à l'avance. A la date t, le prix spot est noté $S_t = F(t, t+1, t+1)$. Le troisième concerne les produits dérivés sur les prix de ces deux premiers contrats. Certaines de ces options font intervenir plusieurs énergies comme le spark spread (gaz et électricité). Dès lors, modéliser conjointement les dynamiques des prix des contrats devient une problèmatique intéressante. De plus, les modèles de prix pour les énergies sont fondamentalement différents de ceux des prix des actions sur les marchés financiers. Par exemple, l'évolution des prix des contrats spots gaz et électricité possèdent des propriétés statistiques particulières: saisonnalité, structures de corrélations forte, stationnarité, pics de prix, queues de distribution épaisses, ... Toutes ces particularités doivent être prises en compte afin de bien valoriser les actifs physiques (Turbine à gaz, stockages, options swing, ...) et les options sur les marchés du gaz et de l'électricité. Ceci est l'une des motivations qui nous a conduit à nous intéresser à la modélisation conjointe des prix spot du gaz et de l'électricité.

De nombreux modèles ont été proposés dans la littérature. Généralement, ils sont basés sur les processus d'Ornstein afin de modéliser la propriété de retour à la moyenne des prix. Des modèles d'équilibre ont été proposés afin de reproduire le phénomène d'offre et de demande dans la formation des prix. Nous nous référons à [7] et [46] pour plus de détails concernant cette approche. Son principal désavantage réside dans la structure d'autocorrélation des prix du gaz et de l'électricité ainsi que la structure de corrélation croisée qui ne sont pas reproduites correctement.

Afin de bien représenter le phénomène de pics de prix, [37] propose une diffusion markovienne avec sauts. Cependant, appliqué aux prix des contrats spot gaz et électricité, ce modèle ne pourrait pas représenter les structures d'autocorrélations et de corrélations croisées des deux énergies.

Une autre classe de dynamiques récemment introduite est celle des modèles multi-facteurs. De nombreux auteurs ont étudié ce type de diffusion. Nous pouvons citer [10], [20], [66], [85] parmi bien d'autres. Le logarithme du prix (on parlera alors de modèle géométrique) ou directement le prix (on parlera alors de modèle arithmétique) est représenté par une somme de processus d'Ornstein. Certains de ces processus sont à sauts purs tandis que d'autres sont des processus avec des variations plus régulières. Par exemple dans [66], un modèle à deux facteurs est étudié. Le log des prix désaisonnalisés $X(t)$ est modélisé par

$$X(t) = Y_1(t) + Y_2(t)$$

où

$$\mathrm{d}Y_i(t) = -\lambda_i Y_i(t)\mathrm{d}t + \mathrm{d}L_i(t), \quad i = 1, 2.$$

Y_1 est un processus d'Ornstein Uhlenbeck (OU), *i.e.* L_1 est un mouvement brownien standard alors que Y_2 est un processus d'Ornstein à sauts discontinues, *i.e.* L_2 est un processus à sauts purs. Dans ce type de modèle, la difficulté réside dans la détection et le filtrage des pics de prix afin d'estimer les paramètres du processus à sauts. Plusieurs méthodes ont été proposées pour résoudre ce problème. Nous nous référons à [66] et [10] pour plus de détails au sujet du processus de calibration associé à ces modèles.

Dans ce troisième chapitre, afin de représenter l'ensemble des propriétés statistiques observées par les prix des contrats spot du gaz et de l'électricité, nous proposons une nouvelle classe de modèles multi-facteurs (géométrique et arithmétique) basée sur des processus d'Ornstein à volatilité locale paramétrique. La principale motivation de ce travail réside dans l'article [15]. Dans celui-ci, les auteurs montrent qu'il est possible à partir d'une fonction d'autocorrélation décroissante en exponentielle et une loi marginale données de construire un processus de diffusion avec drift linéaire et coefficient de diffusion paramétrique reproduisant la structure de corrélation demandée et dont la loi stationnaire correspond à la loi marginale. Afin de reproduire des pics de prix nous utilisons ce type de processus. Cela nous permet d'éviter d'introduire des processus à sauts. Notre processus de calibration est basé sur des méthodes statistiques robustes et rapides ce qui est un avantage considérable comparé à la classe des diffusions à sauts. Notre modèle nous permet de reproduire les structures d'autocorrélation de chaque énergie ainsi que la structure de corrélation croisée entre les prix spot gaz et électricité. Nous avons choisi de réaliser cette étude dans le cas du modèle géométrique afin de garantir la positivité des prix.

Les principales propriétés statistiques des prix spot du gaz et de l'électricité sont:

- la première et sans doute la plus partagée par les autres matières premières est la saisonnalité. Elle est généralement modélisée par un trend (fonction affine du temps) et une somme de sinusoïdes. Le log des prix de chaque énergie

fluctuent autour des deux fonctions

$$\log g(t) = a^g + b^g t + \sum_{k=1}^{m} c_k^g \cos\left(\frac{2\pi t}{l_k}\right) + d_k^g \sin\left(\frac{2\pi t}{l_k}\right),$$
$$\log e(t) = a^e + b^e t + \sum_{k=1}^{m} c_k^e \cos\left(\frac{2\pi t}{l_k}\right) + d_k^e \sin\left(\frac{2\pi t}{l_k}\right),$$

où $l_k = \lfloor 252/k \rfloor$, $k = 1, ..., m$, et $\lfloor x \rfloor$ est la partie entière de x. Pour $m = 2$, nous prenons en compte une saisonnalité annuelle et semestrielle. Tous les coefficients sont calibrés par une méthode des moindre carrés ordinaires.

- l'électricité n'étant pas une énergie stockable cela entraîne la présence de pics de prix. Le gaz peut être stocké mais pas à moindre coût, par conséquent ce dernier partage la propriété de pics de prix avec l'électricité comme nous le montre les historiques de prix suivants. Cela a pour conséquence directe que

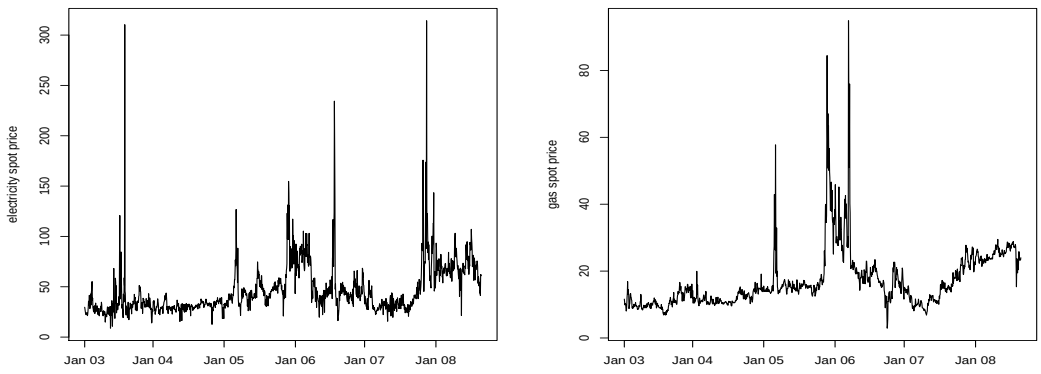


Figure 1.1: Prix des contrats spot électricité sur le marché de Powernext (sur la gauche) et du gaz au NBP (sur la droite) pour la période du 14 Janvier 2003 jusqu'au 20 Août 2008.

les distributions marginales des prix sont à queues épaisses et disymétriques.

- Une fois la saisonnalité enlevée, nous avons remarqué que les séries des log-prix désaisonnalisés sont stationnaires (phénomène de retour à la moyenne) et liées entre elles en niveau. Il existe une dépendance long terme entre les deux énergies. Ces deux propriétés sont observables au niveau des fonctions d'autocorrélation et de corrélation croisée. Elles peuvent être représentées par une somme de fonctions exponentielles décroissantes avec différentes vitesses, *i.e.* que nous choisissons la modélisation suivante pour $\tau > 0$

$$\rho^g(\tau) = \mathrm{Corr}\left(Y^g(t+\tau), Y^g(t)\right) = \phi_1^g e^{-\lambda_1^g \tau} + (1-\phi_1^g) e^{-\lambda_2^g \tau},$$
$$\rho^e(\tau) = \mathrm{Corr}\left(Y^e(t+\tau), Y^e(t)\right) = \phi_1^e e^{-\lambda_1^e \tau} + (1-\phi_1^e) e^{-\lambda_2^e \tau},$$
$$\rho^{g,e}(\tau) = \mathrm{Corr}\left(Y^g(t+\tau), Y^e(t)\right) = \phi^{g,e} e^{-\lambda^{g,e} \tau}.$$

Les vitesses de décroissance λ^g, λ^e sont les vitesses rapides de retour à la moyenne, celles qui permettront après un pic de prix de revenir autour de la moyenne alors que $\lambda^{g,e}$ est la vitesse lente et correspond à la vitesse de retour à la moyenne du processus à variations normales ou d'équilibre. Notons que dans notre modélisation nous avons choisi la même vitesse lente $\lambda^{g,e}$ dans les fonctions d'autocorrélation du gaz et de l'électricité. Ces coefficients sont calibrés en utilisant la méthode des moindres carrés ordinaire.

Afin de reproduire ces propriétés statistiques grâce à notre modèle, nous étudions la construction d'un processus de diffusion X à valeurs dans l'intervalle (l, r) solution d'une EDS dont la loi invariante f et la fonction d'autocorrélation sont connues et données.

Proposition 1.3.1. *Soit $b : x \in (l, r) \mapsto -\lambda(x - \mu)$. On suppose que la densité f a pour moyenne μ et une variance finie. Alors il existe une unique fonction σ définie par*

$$\forall x \in (l, r), \quad \sigma(x) = \sqrt{\frac{\int_l^x 2\lambda(\mu - y) f(y) \mathrm{d}y}{f(x)}},$$

tel que l'EDS

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \quad X_0 \in (l, r), \tag{$E_{b,\sigma}$}$$

a une unique solution X, qui est un processus de diffusion ergodique de distribution stationnaire ν satisfaisant $\nu(\mathrm{d}x) = f(x)\mathrm{d}x$, et de fonction d'autocorrélation

$$\forall t, \tau \geq 0, \quad \mathrm{Corr}\left(X_{t+\tau}, X_t\right) = e^{-\lambda\tau}.$$

La fonction σ est calculable de manière explicite pour la majorité des distributions utilisées en pratique. Cependant, pour certaines d'entre elles, aucune formule explicite existe. Dans ce cas, une approximation par point selle est envisageable. Celle-ci est basée sur la fonction génératrice des moments $M(t)$ et son logarithme $\kappa(t) = \log(M(t))$. L'approximation par point selle d'ordre 1 de la densité f est donnée par la densité $\hat{f}$

$$\forall x \in (l, r), \quad \hat{f}(x) = \left(2\pi\kappa''(\hat{t}_x)\right)^{-1/2} e^{-\left(\hat{t}_x x - \kappa(\hat{t}_x)\right)},$$

où $t = \hat{t}_x$ est la solution (unique) de l'équation $\kappa'(t) = x$. En général, pour les distributions où le coefficient de diffusion n'est pas explicite dans la proposition 1.3.1, $\hat{t}_x$ est connue et une formule "simple" existe.

Une comparaison entre $\hat{f}$ et la mesure de vitesse de $(E_{b,\sigma})$, nous suggère la relation $t = \frac{-2b}{\sigma^2}$. Cela nous amène alors à considérer la solution de l'EDS $(E_{b,\sigma})$ dont le drift est $b : x \in (l, r) \mapsto -\lambda(x - \mu)$ et le coefficient de diffusion est défini par

$$\forall x \in (l, r), \quad \sigma(x) = \sqrt{\frac{-2b(x)}{\hat{t}(x)}}.$$

Celle-ci possède une unique solution X ergodique de distribution stationnaire $\tilde{f}(x) = \frac{c}{\sigma^2(x)} e^{-\left(x\hat{t}(x) - \kappa(\hat{t}(x))\right)}$ où c est une constante de normalisation. Nous étudions l'exemple

de la loi NIG (Normal Inverse Gaussian) qui fait partie de la famille des distributions hyperboliques généralisées. La densité de la loi NIG est définie par

$$f(x) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+(x-l)^2}\right)}{\pi\sqrt{\delta^2+(x-l)^2}} \times e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-l)}, \quad x \in \mathbb{R},$$

où $\beta \in \mathbb{R}$, $\alpha > |\beta|$, $\delta > 0$, $l \in \mathbb{R}$ et K_1 est la fonction de Bessel modifiée d'ordre 3 et d'indexe 1. Cette distribution permet de générer de grandes valeurs avec une probabilité élevée. Nous nous baserons sur cette distribution afin de générer des pics de prix. La fonction $\hat{t}$ est connue explicitement et l'EDS ayant pour densité $\tilde{f}$ est entièrement spécifié par les deux fonctions

$$\forall x \in \mathbb{R}, \quad b(x) = -\lambda\,(x-\mu) \quad \text{et} \quad \sigma^2(x;\theta) = \frac{2\lambda\sqrt{\delta^2+(x-l)^2}\,(x-\mu)}{\alpha\,(x-l) - \beta\sqrt{\delta^2+(x-l)^2}}, \quad (1.52)$$

où $\mu = l + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}}$ et $\theta := (\alpha, \beta, \delta, l)$ représente les quatre paramètres de la loi.

Le modèle que nous proposons est basé sur une somme de diffusions définies par $(E_{b,\sigma})$ avec des choix judicieux des fonctions b et σ. Afin de bien reproduire les fonctions d'autocorrélation et de corrélation croisée des séries log-prix désaisonnalisés Y^g (pour le gaz) et Y^e (pour l'électricité), nous étudions la dynamique à deux facteurs suivante

$$Y_t^g = X_t^g + Z_t, \quad \text{et} \quad Y_t^e = X_t^e + Z_t,$$

où $(Z_t)_{t\geq 0}$, $(X_t^g)_{t\geq 0}$ et $(X_t^e)_{t\geq 0}$ sont des processus mutuellement indépendants définies comme suit:

- Le diffusion Z est le processus d'équilibre entre les deux énergies avec une vitesse de retour à la moyenne lente $\lambda_z = \lambda_2^g = \lambda_2^e$. Ce processus représente les variations normales et nous choisissons pour celui-ci un processus d'Ornstein-Uhlenbeck

  $$\mathrm{d}Z_t = -\lambda_z Z_t \mathrm{d}t + \sigma_z \mathrm{d}W_t^z,$$

  avec $\lambda_z > 0$ et $\sigma_z > 0$.

- Les diffusions X^g et X^e représentent les processus responsables des pics de prix pour chaque énergie. Nous les modélisons par des processus de loi invariante NIG et de vitesses de retour à la moyenne rapides $\lambda_g = \lambda_1^g > 0$ and $\lambda_e = \lambda_2^e > 0$, *i.e.*

  $$\mathrm{d}X_t^j = -\lambda_j\left(X_t^j - \mu_j\right)\mathrm{d}t + \sigma_j(X_t^j;\theta_j)\mathrm{d}W_t^j, \ j = g, e,$$

  où $\sigma_j(.;\theta)$ est définie par (1.52).

Ensuite, nous proposons un processus de calibration adapté à ce modèle et basé sur des outils statistiques rapides et robustes. Sans rentrer dans le détail, les grandes étapes sont les suivantes:

- **Etape 1**: On désaisonnalise le log des prix par régression sur les fonctions $\log g$ et $\log e$ à l'aide de la méthode des moindre carrés ordirnaire.

- **Etape 2**: Ensuite, on calibre les coefficients apparaissant dans les fonctions d'autocorrélation et de corrélations croisées par la méthode des moindre carrés encore une fois.

- **Etape 3**: Puis, pour finir, on estime les coefficients θ^g et θ^e en utilisant la méthode du maximum de vraisemblance tronquée d'ordre $m \geq 0$. Celle-ci consiste à approximer la vraisemblance de Y^g et Y^e dont on ne peut pas obtenir une formule explicite par la vraisemblance tronquée d'ordre $m \geq 0$, *i.e.* celle qui consiste à tenir compte à chaque date des m données passées. Dans [5], il est montré qu'un tel estimateur est fortement consistant et vérifie asymptotiquement un TCL. Dans notre méthode de calibration, nous étudions les estimateurs de θ_g et θ_e d'ordre 0 et 1. L'estimateur d'ordre 0 est plus robuste et fournit de meilleurs résultats sur les simulations effectuées.

Dans la dernière partie, nous proposons des simulations du modèle calibré, ainsi qu'une comparaison de trois différents modèles appliqués à la valorisation (calcul d'espérance sous la probabilité historique) et au contrôle du risque (calcul de la VaR_α) d'un portefeuille modélisant de manière simple le comportement d'une turbine à gaz. Cette comparaison a pour but de mettre en évidence l'importance de la prise en compte des phénomènes de pics de prix et de lien long terme entre les prix spot du gaz et de l'électricité.

Part I

Computation of VaR and CVaR using stochastic approximations

Chapter 2

Computation of VaR and CVaR using stochastic approximations and unconstrained importance sampling

This work appeared in *Monte Carlo Methods and Applications*.

Joint paper with O. Bardou and G. Pagès.

Abstract: Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) are two risk measures which are widely used in the practice of risk management. This paper deals with the problem of estimating both VaR and CVaR using stochastic approximation (with decreasing steps): we propose a first Robbins-Monro (RM) procedure based on Rockafellar-Uryasev's identity for the CVaR. Convergence rate of this algorithm to its target satisfies a Gaussian Central Limit Theorem. As a second step, in order to speed up the initial procedure, we propose a recursive and adaptive importance sampling (IS) procedure which induces a significant variance reduction of both VaR and CVaR procedures. This idea, which has been investigated by many authors, follows a new approach introduced in [62]. Finally, to speed up the initialization phase of the IS algorithm, we replace the original confidence level of the VaR by a slowly moving risk level. We prove that the weak convergence rate of the resulting procedure is ruled by a Central Limit Theorem with minimal variance and its efficiency is illustrated on several typical energy portfolios.

Keywords: VaR, CVaR, Stochastic Approximation, Robbins-Monro algorithm, Importance Sampling, Girsanov Transform.

2.1 Introduction

Following financial institutions, energy companies are developing a risk management framework to face the new price and volatility risks associated to the growth of energy markets. Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are certainly the best known and the most common risk measures used in this context, especially for the evaluation of extreme losses potentially faced by traders. Naturally related to rare events, the estimation of these risk measures is a numerical challenge. The Monte Carlo method, which is often the only available numerical device in such a general framework, must always be associated to efficient reduction variances techniques to encompass its slow convergence rate. In some specific cases, Gaussian approximations can lead to semi-closed form estimators. But, if these approximations can be of some interest when considering the *yield* of a portfolio, they turn out to be useless when estimating e.g. the VaR on the EBITDA (Earnings Before Interest, Taxes, Depreciation, and Amortization) of a huge portfolio as it is often the case in the energy sector.

In this article, we introduce an alternative estimation method to estmate both VaR and CVaR, relying on the use of recursive stochastic algorithms. By definition, the VaR at level $\alpha \in (0,1)$ (VaR_α) of a given portfolio is the lowest amount not exceeded by the loss with probability α (usually $\alpha \geq 95\%$). The Conditional Value-at-Risk at level α (CVaR_α) is the conditional expectation of the portfolio losses beyond the VaR_α level. Compared to VaR, CVaR is known to have better properties. It is a coherent risk measure in the sense of Artzner, Delbaen, Eber and Heath, see [3].

The most commonly used method to compute VaR is the inversion of the simulated empirical loss distribution function using Monte Carlo or historical simulation tools. The historical simulation method usually assumes that the asset returns in the future are independent and identically distributed, having the same distribution as they had in the past. Over a time interval $[t,T]$, the loss is defined by $L := V(S_t,t) - V(S_t + \Delta S, T)$, where S_t denotes the market price vector observed at time t, $\Delta S = S_T - S_t$ the variation of S over the time interval $[t,T]$ -which can be calculated using historical data- and $V(S_t,t)$ the portfolio value at time t. The distribution of this loss L can be computed with the corresponding VaR at a given probability level by the inversion of the empirical function method. However, when the market price dynamics follow a general diffusion process solution of a stochastic differential equation (SDE), the assumption of asset returns independence is no longer available.

To circumvent this problem, Monte Carlo simulation tools are generally used. Another widely used method relies on a linear (Normal approximation) or quadratic expansion (Delta-Gamma approximation) and assume a joint normal (or log-normal) distribution for ΔS. The Normal approximation method gives L a normal distribution, thus the computation of the VaR_α is straightforward. However, when there is a non-linear dependence between the portfolio value and the prices of the underlying assets (think of a portfolio with options) such approximation is no longer acceptable. The Delta-Gamma approximation tries to capture some non linearity by adding a quadratic term in the loss expansion. Then, it is possible to find the distribution of the resulting approximation in order to obtain an approximation of the VaR. For more details about these methods, we refer to [17], [21], [39], [40] and [80]. Such

approximations are no longer acceptable when considering portfolios with long maturity ($T - t = 1$ year up to 10 years) or when the loss is a functional of a general path-dependent SDE.

In the context of hedging or optimizing a portfolio of financial instruments to reduce the CVaR, it is shown in [77] that it is possible to compute both VaR and CVaR (actually calculate VaR and optimize CVaR) by solving a convex optimization problem with a linear programming approach. It consists in generating loss scenarios and then in introducing constraints in the linear programming problem. Although they address a different problem, this method can be used to compute both VaR and CVaR. The advantage of such a method is that it is possible to estimate both VaR and CVaR simultaneously without assuming that the market prices have a specified distribution (e.g. normal, log-normal, ...). The main drawback is that the dimension (number of constraints) of the linear programming problem to be solved is equal to the number of simulated scenarios. In our approach, we are no limited by the number of generated sample paths used in the procedure.

The idea to compute both VaR and CVaR with one procedure comes from the fact that they are strongly linked as they appear as the solutions and the value of the same convex optimisation problem (see Proposition 2.1) as pointed out [77]. Moreover both the objective function of the minimization problem and its gradient read as an expectation. This leads us to define consistent and asymptotically normal estimators of both quantities as the limit of a global Robbins-Monro (RM) procedure. Consequently, we are no longer constrained by the number of samples paths used in the estimation.

A significant advantage of this recursive approach, especially in regard to the inversion of the empirical function method is that we only estimate the quantities of interest and not the whole inverse of the distribution function. Furthermore, we do not need to make approximations of the loss or of the convex optimization problem to be solved. Moreover, the implementation of the algorithm is straightforward. However to make it really efficient we need to modify it owing to the fact that VaR and CVaR computation is closely related to the simulation of rare events. That is why as a necessary improvement, we introduce a (recursive and adaptive) variance reduction method based on an importance sampling (IS) paradigm.

Let us be a bit more specific. Basically in this kind of problem we are interested in events that are observed with a very low probability (usually less that 5%, 1% or even 0.1%) so that we obtain few significant replications to update our estimates. Actually, interesting losses are those that exceed the VaR, *i.e.* the ones that are "in the tail" of the loss distribution. Thus in order to compute more accurate estimates of both quantities of interest, it is necessary to generate more samples in the tail of L, the area of interest. A general tool used in this situation is IS.

The basic principle of IS is to modify the distribution of L by an equivalent change of measure to obtain more "interesting" samples that will lead to better estimates of the VaR and CVaR. The main issue of IS is to find a right change of measure (among a parameterized family) that will induce a significant variance reduction. In [39] and [38], a change of measure based on a large deviation upper bound is proposed to estimate the loss probability $\mathbb{P}(L > x)$ for several values of x. Then, it is possible to estimate the VaR by interpolating between the estimated loss probabilities.

Although this approach provides an asymptotically optimal IS distribution, it is strongly based on the fact that the Delta-Gamma approximation holds exactly and relies on the assumption that, conditionally to the past data, market moves are normally distributed. Moreover, as shown in [41], importance sampling estimators based on a large deviations change of measure can have variance that increases with the rarity of the event, and even infinite variance. In [25], the VaR_α is estimated by using a quantile estimator based on the inversion of the empirical weighted function and combined with Robbins-Monro (RM) algorithm with repeated projection devised to produce the optimal measure change for IS purpose. This kind of IS algorithm is known to converge toward the optimal importance sampling parameter only after a (long) stabilization phase and provided that the compact sets have been appropriately specified. By contrast, our parameters are optimized by an adaptive unconstrained (*i.e.* without projections) RM algorithm naturally combined with our VaR-CVaR procedure.

One major issue that arises when combining the VaR-CVaR algorithm with the recursive IS procedure is to ensure that the IS parameters do move appropriately toward the critical risk area. They may remain stuck at the very beginning of the IS procedure. To circumvent this problem, we make the confidence level slowly increase from a low level (say 50%) to α by introducing a deterministic sequence $(\alpha_n)_{n\geq 0}$ of confidence level that converges toward α. This kind of incremental threshold increase has been proposed previously [55] in a different framework (use of cross entropy in rare event simulation). It speeds up the initialization phase of the IS algorithm and consequently improves the variance reduction. Thus, we can truly experiment asymptotic convergence results in practice.

The paper is organized as follows. In the next section, we present some theoretical results about VaR and CVaR. We introduce the VaR-CVaR stochastic algorithm in its first and naive version and study its convergence rate. We also introduce some background about IS using stochastic approximation algorithm. Section 3 is devoted to the design of an optimal procedure using an adaptive variance reduction procedure. We present how it modifies the asymptotic variance of our first CLT. In Section 4 we provide some extensions to the exponential change of measure and to deal with the case of infinite dimensional setting. Section 5 is dedicated to numerical examples. We propose several portfolios of options on several assets in order to challenge the algorithm and display variance reduction factors obtained using the IS procedure. To prevent the freezing of the algorithm during the first iterations of the IS procedure, we also consider a deterministic moving risk level α_n which replace α to speed up the initialization phase and improve the reduction of variance. We prove theoretically that modifying in this way the algorithm doesn't change the previous CLT and fasten the convergence.

Notations: • $|.|$ will denote the canonical Euclidean norm on $\mathbb{R}^d$ and $\langle ., . \rangle$ will denote the canonical inner product.

• $\xrightarrow{\mathcal{L}}$ will denote the convergence in distribution and $\xrightarrow{a.s.}$ will denote the almost sure convergence.

• $x_+ := \max(0, x)$ will denote the positive part function.

2.2 VaR, CVaR using stochastic approximation and some background on recursive IS

It is rather natural to consider that the loss of the portfolio over the considered time horizon can be written as a function of a structural finite dimensional random vector, *i.e.* $L = \varphi(X)$, where X is a $\mathbb{R}^d$-valued random vector defined on the probability space $(\Omega, A, \mathbb{P})$ and $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a Borel function. φ is the function representing the composition of the portfolio which remains fixed and X is a structural random vector used to model the market prices over the time interval; therefore we do not need to specify the dynamics of the market prices and only rely on the fact that it is possible to sample from the distribution of X. For instance, in a Black-Scholes framework, X is a Gaussian vector and φ can be a portfolio of vanilla options. In more sophisticated models or portfolio, X can be a vector of Brownian increments related to the Euler scheme of a diffusion. The VaR at level $\alpha \in (0,1)$ is the lowest α-quantile of the distribution $\varphi(X)$ *i.e.*:

$$\mathrm{VaR}_\alpha(\varphi(X)) := \inf\left\{\xi \mid \mathbb{P}\left(\varphi(X) \leq \xi\right) \geq \alpha\right\}.$$

Since $\lim_{\xi\to+\infty} \mathbb{P}\left(\varphi(X) \leq \xi\right) = 1$, we have $\left\{\xi \mid \mathbb{P}\left(\varphi(X) \leq \xi\right) \geq \alpha\right\} \neq \varnothing$. Moreover, we have $\lim_{\xi\to-\infty} \mathbb{P}\left(\varphi(X) \leq \xi\right) = 0$, which implies that $\left\{\xi \mid \mathbb{P}\left(\varphi(X) \leq \xi\right) \geq \alpha\right\}$ is bounded from below so that the VaR always exists. We assume that the distribution function of $\varphi(X)$ is continuous (*i.e.* without atoms) so that the VaR is the lowest solution of the equation:

$$\mathbb{P}\left(\varphi(X) \leq \xi\right) = \alpha.$$

Three values of α are commonly considered: 0.95, 0.99, 0.995 so that it is usually close to 1 and the tail of interest has probability $1-\alpha$. If the distribution function is (strictly) increasing, the solution of the above equation is unique, otherwise, there may be more than one solution. In fact, in what follows, we will consider that *any* solution of the previous equation is the VaR. Another risk measure generally used to provide information about the tail of the distribution of $\varphi(X)$ is the *Conditional Value-at-Risk* (CVaR) (at level α). As soon as $\varphi(X) \in L^1(\mathbb{P})$, it is defined by:

$$\mathrm{CVaR}_\alpha(\varphi(X)) := \mathbb{E}\left[\varphi(X) | \varphi(X) \geq \mathrm{VaR}_\alpha(\varphi(X))\right].$$

The CVaR of $\varphi(X)$ is simply the conditional expectation of $\varphi(X)$ given that it lies inside the critical risk area. To capture more information on the conditional distribution of $\varphi(X)$, it seems natural to consider more general risk measures like for example the conditional variance. In a more general framework we can be interested in estimating the Ψ-*Conditional Value at Risk* (Ψ-CVaR) (at level α) where $\Psi : \mathbb{R} \to \mathbb{R}$ is a continuous function. As soon as $\Psi(\varphi(X)) \in L^1(\mathbb{P})$, it is defined by:

$$\Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X)) := \mathbb{E}\left[\Psi(\varphi(X)) | \varphi(X) \geq \mathrm{VaR}_\alpha(\varphi(X))\right]. \tag{2.1}$$

When $\Psi \equiv Id$ and $\varphi(X) \in L^1(\mathbb{P})$, (2.1) is the regular CVaR of $\varphi(X)$. When $\Psi \equiv x \mapsto x^2$, equation (2.1) is but the conditional quadratic norm of $\varphi(X)$.

2.2.1 Representation of VaR and Ψ-CVaR as expectations

The idea to devise a stochastic approximation procedure to compute VaR and CVaR, and more generally the Ψ-CVaR, comes from the fact that these two quantities are solutions of a convex optimization problem whose value function can be represented as an expectation as pointed out by Rockafellar and Uryasev in [77].

Proposition 2.2.1. *Let V and V_Ψ be the functions defined by:*

$$V(\xi) = \mathbb{E}\left[v(\xi, X)\right] \quad \textit{and} \quad V_\Psi(\xi) = \mathbb{E}\left[w(\xi, X)\right] \tag{2.2}$$

where

$$v(\xi, x) := \xi + \frac{1}{1-\alpha}(\varphi(X) - \xi)_+ \quad \textit{and} \quad w(\xi, x) := \xi + \frac{1}{1-\alpha}(\Psi(\varphi(x)) - \xi)\mathbf{1}_{\{\varphi(x)\geq\xi\}}. \tag{2.3}$$

Suppose that the distribution function of $\varphi(X)$ is continuous and that $\varphi(X) \in L^1(\mathbb{P})$. Then, the function V is convex, differentiable and the $\mathrm{VaR}_\alpha(\varphi(X))$ is any point of the set:

$$\arg\min V = \{\xi \in \mathbb{R} \mid V'(\xi) = 0\} = \{\xi \mid \mathbb{P}(\varphi(X) \leq \xi) = \alpha\},$$

where V' is the derivative of V defined for every $\xi \in \mathbb{R}$ by

$$V'(\xi) = \mathbb{E}\left[\frac{\partial v}{\partial \xi}(\xi, X)\right]. \tag{2.4}$$

Furthermore,

$$\mathrm{CVaR}_\alpha(\varphi(X)) = \min_{\xi\in\mathbb{R}} V(\xi)$$

and, if $\Psi\left(\varphi(X)\right) \in L^1(\mathbb{P})$, for every $\xi^* \in \arg\min V$ (i.e., ξ^* is a $\mathrm{VaR}_\alpha(\varphi(X))$)

$$\Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X)) = V_\Psi(\xi^*).$$

Proof. Since the functions $\xi \mapsto (\varphi(x) - \xi)_+$, $x \in \mathbb{R}^d$, are convex, the function V is convex. $\mathbb{P}(dw)$-*a.s.*, $\frac{\partial v}{\partial \xi}(\xi, X(w))$ exists at every $\xi \in \mathbb{R}$ and

$$\mathbb{P}(dw)\text{-}a.s., \quad \left|\frac{\partial v}{\partial \xi}(\xi, X(w))\right| \leq 1 \vee \frac{\alpha}{1-\alpha}.$$

Thanks to Lebesgue Dominated Convergence Theorem, one can interchange differentiation and expectation, so that V is differentiable with derivative $V'(\xi) = 1 - \frac{1}{1-\alpha}\mathbb{P}(\varphi(X) > \xi)$ and reaches its absolute minimum at any ξ^* satisfying $\mathbb{P}(\varphi(X) > \xi^*) = 1 - \alpha$ *i.e.* $\mathbb{P}(\varphi(X) \leq \xi^*) = \alpha$.
Moreover, it is clear that:

$$\begin{aligned} V(\xi^*) &= \xi^* + \frac{\mathbb{E}[(\varphi(X) - \xi^*)_+]}{\mathbb{P}(\varphi(X) > \xi^*)} \\ &= \frac{\xi^*\mathbb{E}[\mathbf{1}_{\varphi(X)>\xi^*}] + \mathbb{E}[(\varphi(X) - \xi^*)_+]}{\mathbb{P}(\varphi(X) > \xi^*)} \\ &= \mathbb{E}\left[\varphi(X)|\varphi(X) > \xi^*\right] \end{aligned}$$

and, in the same way, $V_\Psi(\xi^*) = \Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X))$. This completes the proof. □

Remark: Actually, one could consider a more general framework by including any risk measure defined by an integral representation with respect to X:

$$\mathbb{E}[\Lambda(\xi^*, X)]$$

where Λ is a (computable) Borel function.

2.2.2 Stochastic gradient and its adaptive companion procedure: a first naive approach

The above representation (2.4) naturally yields a stochastic gradient procedure derived from the convex Lyapunov function V which will (hopefully) converge toward $\xi^* := \mathrm{VaR}_\alpha(\varphi(X))$. Then, a recursive companion procedure based on (2.2) can be easily devised having $C^* := \Psi\text{-CVaR}_\alpha(\varphi(X))$ as target. There is no reason to believe that this first version can do better than the empirical quantile estimate. But, it is a necessary phase in order to understand how our recursive IS algorithm (to be devised further on) can be combined with this first procedure.
First we set

$$H_1(\xi, x) := \frac{\partial v}{\partial \xi}(\xi, x) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(x)\geq\xi\}}, \tag{2.5}$$

so that,

$$V'(\xi) = \mathbb{E}\left[H_1\left(\xi, X\right)\right].$$

Since we are looking for ξ for which $\mathbb{E}\left[H_1(\xi, X)\right] = 0$, we implement a stochastic gradient descent derived from the Lyapunov function V to approximate $\xi^* := VaR_\alpha(\varphi(X))$, *i.e.*, we use the RM algorithm:

$$\xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), n \geq 1, \quad \xi_0 \in L^1(\mathbb{P}), \tag{2.6}$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence of random variables with the same distribution as X, independent of ξ_0, with $\mathbb{E}[|\xi_0|] < +\infty$ and $(\gamma_n)_{n\geq 1}$ is a deterministic step sequence (decreasing to 0) satisfying:

$$\sum_{n\geq 1}\gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1}\gamma_n^2 < +\infty. \tag{A1}$$

In order to derive the *a.s.* convergence of (2.6) we introduce the following additional assumption on the distributions of $\varphi(X)$ and $\Psi(\varphi(X))$. Let $a > 0$,

$$\varphi(X) \text{ has a continuous distribution function and } \Psi(\varphi(X)) \in L^{2a}(\mathbb{P}). \tag{$A2)_a$}$$

Actually, Equation (2.6) can be seen either as a regular RM procedure with mean function V' since it is increasing (see e.g. [23] p.50 and p.66) or as a recursive gradient descent procedure derived from the Lyapunov function V. Both settings yield the *a.s.* convergence toward its target ξ^*. To establish the *a.s.* convergence of $(\xi_n)_{n\geq 1}$ under assumptions $(A1)$ and $(A2)_1$ (and of our different RM algorithms), we will rely on the following theorem. For a proof of this slight extension of Robbins-Monro Theorem we refer to the appendix.

Theorem 2.2.2. *(Robbins-Monro Theorem (variant)). Let $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function and X be an $\mathbb{R}^d$-valued random vector such that $\mathbb{E}[|H(z,X)|] < \infty$ for every $z \in \mathbb{R}^d$. Then set*

$$\forall z \in \mathbb{R}^d, \quad h(z) = \mathbb{E}[H(z,X)].$$

Suppose that the function h is continuous and that $\mathcal{T}^* := \{h = 0\}$ satisfies

$$\forall z \in \mathbb{R}^d \ \ | \mathcal{T}^*, \forall z^* \in \mathcal{T}^*, \quad \langle z - z^*, h(z) \rangle > 0. \tag{2.7}$$

Let $(\gamma_n)_{n\geq 1}$ be a deterministic step sequence satisfying condition (A1). Suppose that

$$\forall z \in \mathbb{R}^d, \quad \mathbb{E}[|H(z,X)|^2] \leq C(1 + |z|^2) \tag{2.8}$$

(which implies that $|h(z)| \leq C'(1 + |z|)$).
Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence of random vectors having the distribution of X, let z_0 be a random vector independent of $(X_n)_{n\geq 1}$ satisfying $\mathbb{E}[|z_0|] < \infty$, all defined on the same probability space $(\Omega, A, \mathbb{P})$. Let $\mathcal{F}_n := \sigma(z_0, X_1, ..., X_n)$ and let $(r_n)_{n\geq 1}$ be an $\mathcal{F}_n$-measurable remainder sequence satisfying

$$\sum_n \gamma_n |r_n|^2 < \infty. \tag{2.9}$$

Then, the recursive procedure defined for $n \geq 1$ by

$$Z_n = Z_{n-1} - \gamma_n H(Z_{n-1}, X_n) + \gamma_n r_n,$$

satisfies:

$$\exists\ z_\infty, \quad \text{such that} \ \ Z_n \xrightarrow{a.s.} z_\infty \ \text{and} \ z_\infty \in \mathcal{T}^* \ a.s.$$

The convergence also holds in $L^p(\mathbb{P}), p \in (0,2)$, where $L^p(\mathbb{P})$ denotes the set of all random vectors defined on $(\Omega, A, \mathbb{P})$ such that $\mathbb{E}[|X|^p]^{\frac{1}{p}} < \infty$.

Remark: It is in fact a slight variant (see the appendix for a proof) of the regular RM Theorem since Z_n converges to a random vector having its value in the set $\{h = 0\}$ even if $\{h = 0\}$ is not reduced to a singleton or a finite set. The remainder sequence in the above theorem plays a crucial role when we will (slightly) modify the first IS procedure to improve its efficiency.

The second step concerns procedure for the numerical computation of the Ψ-CVaR$_\alpha$. A naive idea is to compute the function V_Ψ at the point ξ^*:

$$\Psi\text{-CVaR}_\alpha = V_\Psi(\xi^*) = \mathbb{E}[w(\xi^*, X)]$$

using a regular Monte Carlo simulation,

$$\frac{1}{n} \sum_{k=0}^{n-1} w(\xi^*, X_{k+1}). \tag{2.10}$$

However, we first need to get from (2.6) a good approximate of ξ^* and subsequently to use another sample of the distribution X. A natural idea is to devise an adaptive

companion procedure of the above quantile search algorithm by replacing ξ^* in (2.10) by its approximation at step k, namely

$$C_n = \frac{1}{n}\sum_{k=0}^{n-1} w(\xi_k, X_{k+1}),\ n \geq 1,\ C_0 = 0. \tag{2.11}$$

Hence, $(C_n)_{n\geq 0}$ is the sequence of empirical means of the non i.i.d. sequence $(w(\xi_k, X_{k+1}))_{k\geq 1}$, which can be written recursively:

$$C_n = C_{n-1} - \frac{1}{n} H_2\left(\xi_{n-1}, C_{n-1}, X_n\right),\ n \geq 1, \tag{2.12}$$

where $H_2(\xi, c, x) := c - w(\xi, x)$.

At this stage, we are facing two procedures (ξ_n, C_n) with different steps. This may appear not very consistent or at least natural. A second modification to the original Monte Carlo procedure (2.12) consists in considering a general step β_n satisfying condition (A1) instead of $\frac{1}{n}$ (with in mind the possibility to set $\beta_n = \gamma_n$ eventually). This leads to:

$$C_n = C_{n-1} - \beta_n H_2\left(\xi_{n-1}, C_{n-1}, X_n\right),\ n \geq 1. \tag{2.13}$$

In order to prove the *a.s.* convergence of $(C_n)_{n\geq 1}$ toward C^*, we set for convenience $\beta_0 := \sup_{n\geq 1}\beta_n + 1$. Then, one defines recursively a sequence $(\Delta_n)_{n\geq 1}$ by

$$\Delta_{n+1} = \Delta_n \frac{\beta_{n+1}}{\beta_n}\frac{\beta_0}{\beta_0 - \beta_{n+1}},\ n \geq 0,\ \Delta_0 = 1.$$

Elementary computations show by induction that

$$\beta_n = \beta_0 \frac{\Delta_n}{S_n},\ n \geq 0,\ \text{with } S_n = \sum_{k=0}^{n} \Delta_k. \tag{2.14}$$

Furthermore, it follows from (2.14) that for every $n \geq 1$

$$\log(S_n) - \log(S_{n-1}) = -\log\left(1 - \frac{\Delta_n}{S_n}\right) \geq \frac{\Delta_n}{S_n} = \frac{\beta_n}{\beta_0}.$$

Consequently,

$$\log(S_n) \geq \frac{1}{\beta_0}\sum_{k=1}^{n}\beta_k$$

which implies that $\lim_n S_n = +\infty$.
Now using (2.13) and (2.14), one gets for every $n \geq 1$

$$S_n C_n = S_{n-1}C_{n-1} + \Delta_n\left(\Delta N_{n+1} + V_\Psi(\xi_n)\right)$$

where, $\Delta N_n := w(\xi_{n-1}, X_n) - V_\Psi(\xi_{n-1})$, $n \geq 1$, define a martingale increments sequence with respect to the natural filtration of the algorithm $\mathcal{F}_n := \sigma(\xi_0, X_1, \cdots, X_n),\ n \geq 0$. Consequently,

$$C_n = \frac{1}{S_n}\left(\sum_{k=0}^{n-1}\Delta_{k+1}\Delta N_{k+1} + \sum_{k=0}^{n-1}\Delta_{k+1}V_\Psi(\xi_k)\right).$$

The second term in the right hand side of the above equality converges to $V_\Psi(\xi^*) = \Psi\text{-CVaR}_\alpha(\varphi(X))$ owing to the continuity of V_Ψ at ξ^* and Cesaro's Lemma.
The convergence to 0 of the first term will follow from the *a.s.* convergence of the series

$$N_n^\beta := \sum_{k=1}^{n} \beta_k \Delta N_k,\ n \geq 1$$

by the Kronecker Lemma since $\beta_n = \beta_0 \frac{\Delta_n}{S_n}$. The sequence $(N_n^\beta)_{n\geq 1}$ is an $\mathcal{F}_n$-martingale since the ΔN_k's are martingale increments and

$$\mathbb{E}\left[(\Delta N_n)^2 | \mathcal{F}_{n-1}\right] \leq \frac{1}{(1-\alpha)^2} \mathbb{E}\left[\left(\Psi\left(\varphi\left(X\right)\right) - \xi\right)^2\right]_{|\xi=\xi_{n-1}}.$$

The continuity of $\xi \mapsto \mathbb{E}\left[\left(\Psi\left(\varphi\left(X\right)\right) - \xi\right)^2\right]$ at ξ^*, and the *a.s.* convergence of ξ_k toward ξ^* imply that

$$\sup_{n\geq 1} \mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < \infty \quad a.s.$$

Consequently, assumption (A1) implies

$$\langle N^\beta \rangle_\infty = \sum_{n\geq 1} \beta_n^2 \mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < \infty$$

which in term yields the *a.s.* convergence of $(N_n^\beta)_{n\geq 1}$, so that $C_n \xrightarrow{a.s.} \Psi\text{-CVaR}_\alpha\left(\varphi(X)\right)$.
The resulting algorithm reads as for $n \geq 1$:

$$\begin{cases} \xi_n = \xi_{n-1} - \gamma_n H_1\left(\xi_{n-1}, X_n\right), \quad \xi_0 \in L^1(\mathbb{P}), \\ C_n = C_{n-1} - \beta_n H_2\left(\xi_{n-1}, C_{n-1}, X_n\right), \quad C_0 = 0, \end{cases} \tag{2.15}$$

and converges under $(A1)$ and $(A2)_1$.

The question of the joint weak convergence rate of (ξ_n, C_n) is not trivial owing to the coupling of the two procedures. The case of two different step scales refers to the general framework of two-time-scale stochastic approximation algorithms. Several results have been established by Borkar in [16], Konda and Tsitsiklis in [54] but the more relevant in our case are those of Mokkadem and Pelletier in [68]. The weak convergence rate of $(\xi_n)_{n\geq 1}$ is ruled by the CLT for "regular" (single-time scale) stochastic approximation algorithms (we refer to Kushner and Clark in [57], Métivier and Priouret in [13], Duflo in [23] among others). In order to achieve the best asymptotic rate of convergence, one ought to set $\gamma_n = \frac{\gamma_0}{n}$ where the choice of γ_0 depends on the value of the density $f_{\varphi(X)}$ of $\varphi(X)$ at ξ^*, which is unknown. To circumvent the difficulties induced by the specification of γ_0, which are classical in this field, we are led to modify again our algorithm by introducing the averaging principle independently introduced by Ruppert [81] and Polyak [45] and then widely investigated by several authors. It works both with two-time or single-time scale steps and leads to asymptotically efficient procedures, *i.e.*, satisfying a CLT at the optimal rate $\sqrt{n}$ and minimal variance (see also [68]). See also a variant based on a gliding window developed in [61]. Our numerical examples indicate that the averaged one-time-scale procedure provides less variance during the first iterations

than the averaged procedure of the two-time-scale algorithm. Finally, we set $\gamma_n \equiv \beta_n$ in (2.15) so that, the VaR-CVaR algorithm can be written in a more synthetic way by setting $Z_n = (\xi_n, C_n)$ and for $n \geq 1$:

$$Z_n = Z_{n-1} - \gamma_n H(Z_{n-1}, X_n),\ Z_0 = (\xi_0, C_0)\,,\ \xi_0 \in L^1(\mathbb{P}), \tag{2.16}$$

where $H(z,x) := (H_1(\xi,x), H_2(\xi,C,x))$. Throughout the rest of this section, we assume that the distribution $\varphi(X)$ has a positive probability density $f_{\varphi(X)}$ on its support. As a consequence the $\mathrm{VaR}_\alpha(\varphi(X))$ is unique so that the procedure algorithm Z_n converges *a.s.* to its single target $(\mathrm{VaR}_\alpha(\varphi(X)), \Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X)))$. Thus, the Cesaro mean of the procedure

$$\bar{Z}_n := \frac{Z_0 + \cdots + Z_{n-1}}{n},\quad n \geq 1,$$

where Z_n is defined by (2.16), converges *a.s.* to the same target. The Ruppert and Polyak's Averaging Principle says that an appropriate choice of the step yields for free the smallest possible asymptotic variance. We recall below this result (following a version established in [23], see [23] (p.169) for a proof).

Theorem 2.2.3. *(Ruppert and Polyak's Averaging Principle) Suppose that the $\mathbb{R}^d$-sequence $(Z_n)_{n\geq 0}$ is defined recursively by*

$$Z_n = Z_{n-1} - \gamma_n \left(h(Z_{n-1}) + \epsilon_n + r_n\right)$$

where h is a Borel function. Suppose that h is $\mathcal{C}^1$ in the neighborhood of z^* and that $M = Dh(z^*)$ is a uniformly repulsive matrix (all its eigenvalues have positive real parts), and $(\epsilon_n)_{n\geq 1}$ is a random sequence satisfying

$$\exists\, C > 0,\ \textit{such that a.s.}\ \begin{cases} (i)\ \mathbb{E}[\epsilon_{n+1}|\mathcal{F}_n]\mathbf{1}_{\{||Z_n - z^*|| \leq C\}} = 0, \\ (ii)\ \exists b > 2,\ \sup_n \mathbb{E}[||\epsilon_{n+1}||^b|\mathcal{F}_n]\ \mathbf{1}_{\{||Z_n - z^*|| \leq C\}} < +\infty, \\ (iii)\ \mathbb{E}\left[(\gamma_{n-1})^{-1}\,|r_n|^2\ \mathbf{1}_{\{||Z_n - z^*|| \leq C\}}\right] \to 0, \\ (iv)\ \exists\, \Gamma \in \mathcal{S}^+(d,\mathbb{R})\ \textit{such that}\ \mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right] \xrightarrow{a.s.} \Gamma. \end{cases} \tag{2.17}$$

Set $\gamma_n = \frac{\gamma_1}{n^a}$ with $\frac{1}{2} < a < 1$, and

$$\bar{Z}_{n+1} := \frac{Z_0 + ... + Z_n}{n+1} = \bar{Z}_n - \frac{1}{n+1}(\bar{Z}_n - Z_n),\ n \geq 0.$$

Then, on the set of convergence $\{Z_n \to z^*\}$:

$$\sqrt{n}\left(\bar{Z}_n - z^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, M^{-1}\Gamma(M^{-1})^T\right) \qquad \textit{as } n \to +\infty,$$

where $(M^{-1})^T$ denotes the transpose of the matrix M^{-1}.

To apply this theorem to our framework we are led to compute the Cesaro means of both components, namely for $n \geq 1$

$$\begin{cases} \overline{\xi}_n := \frac{1}{n}\sum_{k=1}^n \xi_k = \overline{\xi}_{n-1} - \frac{1}{n}(\overline{\xi}_{n-1} - \xi_n), \\ \overline{C}_n := \frac{1}{n}\sum_{k=1}^n C_k = \overline{C}_{n-1} - \frac{1}{n}(\overline{C}_{n-1} - C_n), \end{cases} \tag{2.18}$$

where (ξ_k, C_k), $k \geq 0$ is defined by (2.16). In the following theorem, we provide the convergence rate of the couple $\bar{Z}_n := \left(\overline{\xi}_n, \overline{C}_n\right)$.

Theorem 2.2.4. *(Convergence rate of the VaR-CVaR procedure). Suppose* $(A2)_a$ *holds for some* $a > 1$*, that the density function* $f_{\varphi(X)}$ *of* $\varphi(X)$ *is continuous and strictly positive at* ξ^**. If the step sequence is* $\gamma_n = \frac{\gamma_1}{n^a}$ *with* $\frac{1}{2} < a < 1$ *and* $\gamma_1 > 0$ *then*

$$\sqrt{n}\left(\bar{Z}_n - z^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma\right) \qquad as\ n \to +\infty$$

where the asymptotic covariance matrix Σ *is given by*

$$\begin{pmatrix} \frac{\alpha(1-\alpha)}{f^2_{\varphi(X)}(\xi^*)} & \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[\left(\Psi(\varphi(X)) - \xi^*\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\ \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[\left(\Psi(\varphi(X)) - \xi^*\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\left(\left(\Psi(\varphi(X)) - \xi^*\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right) \end{pmatrix}. \tag{2.19}$$

Proof. First, the procedure (2.16) can be written as for $n \geq 1$

$$Z_n = Z_{n-1} - \gamma_n\left(h(Z_{n-1}) + \epsilon_n\right), \quad Z_0 = (\xi_0, C_0), \quad \xi_0 \in L^1(\mathbb{P}), \tag{2.20}$$

where $h(z) := \mathbb{E}[H(z,X)] = \left(1 - \frac{1}{1-\alpha}\mathbb{P}\left(\varphi(X) \geq \xi\right), C - \mathbb{E}[w(\xi,X)]\right)$ and $\epsilon_n := (\Delta M_n, \Delta N_n)$, $n \geq 1$, denotes the $\mathcal{F}_n$-adapted martingale increment sequence with

$$\Delta M_n := \frac{1}{1-\alpha}\left(\mathbb{P}\left(\varphi(X) \geq \xi\right)_{|\xi=\xi_{n-1}} - \mathbf{1}_{\{\varphi(X_n)\geq\xi_{n-1}\}}\right).$$

Owing to Assumption $(A2)_a$, the differentiability of Ψ at ξ^* and Lebesgue's differentiation Theorem, one can interchange expectation and derivation, so that the function h is differentiable at $z^* = (\xi^*, C^*)$ and

$$h'(z^*) = M := \begin{pmatrix} \frac{1}{1-\alpha}f_{\varphi(X)}(\xi^*) & 0 \\ \mathbb{E}\left[\left(\frac{\partial}{\partial\xi}w(\xi,X)\right)_{|\xi=\xi^*}\right] & 1 \end{pmatrix}. \tag{2.21}$$

Now, owing to Lebesgue Dominated Convergence Theorem, $\mathbb{E}\left[\left(\frac{\partial}{\partial\xi}w(\xi,X)\right)_{|\xi=\xi^*}\right] = \left(1 - \frac{1}{1-\alpha}\mathbb{P}(\varphi(X) \geq \xi^*)\right) = 0$, so that, $M = \begin{pmatrix} \frac{1}{1-\alpha}f_{\varphi(X)}(\xi^*) & 0 \\ 0 & 1 \end{pmatrix}$ is diagonal and M is uniformly repulsive.
To apply Theorem 2.3, we need to check assumptions (i)-(iv) of (2.17).
Let $A > 0$. First note that

$$\mathbb{E}\left[\Delta M^{2a}_{n+1}|\mathcal{F}_n\right]\mathbf{1}_{\{|Z_n - z^*|\leq A\}} \leq \left(\frac{1}{1-\alpha}\right)^{2a} 2^{2a} < +\infty.$$

Thanks to Assumption $(A2)_a$, there exists $C_{\alpha,\Psi} > 0$ such that

$$\mathbb{E}\left[\Delta N^{2a}_{n+1}|\mathcal{F}_n\right]\ \mathbf{1}_{\{||Z_n - z^*||\leq A\}} \leq C_{\alpha,\Psi}\left(1 + \mathbb{E}\left[\Psi(\varphi(X))^{2a}\right]\right) < +\infty.$$

Consequently, (ii) of (2.17) holds true with $b = 2a > 2$ since

$$\sup_{n\geq 0}\mathbb{E}\left[|\epsilon_{n+1}|^{2a}|\mathcal{F}_n\right]\mathbf{1}_{\{|Z_n - z^*|\leq A\}} < +\infty.$$

It remains to check (iv) for some positive definite symmetric matrix Γ. The dominated convergence theorem implies that

$$\begin{aligned}
\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{1,1}|\mathcal{F}_n\right] &= \left(\frac{1}{1-\alpha}\right)^2\left(\mathbb{E}\left[\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]_{|\xi=\xi_n} - \mathbb{E}\left[\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]^2_{|\xi=\xi_n}\right) \\
&\xrightarrow{a.s.} \frac{\alpha}{1-\alpha}, \\
\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{1,2}|\mathcal{F}_n\right] &= \mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{2,1}|\mathcal{F}_n\right] \\
&= \left(\frac{1}{1-\alpha}\right)^2\mathbb{E}\left[(\Psi(\varphi(X))-\xi)\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]_{|\xi=\xi_n} \\
&\qquad\qquad \times\left(1-\mathbb{E}\left[\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]_{|\xi=\xi_n}\right) \\
&\xrightarrow{a.s.} \frac{\alpha}{(1-\alpha)^2}\mathbb{E}\left[(\Psi(\varphi(X))-\xi^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right],
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{2,2}|\mathcal{F}_n\right] &= \mathbb{E}\left[(\Delta N_{n+1})^2|\mathcal{F}_n\right] \\
&= \frac{1}{(1-\alpha)^2}\Big(\mathbb{E}\left[(\Psi(\varphi(X_{n+1}))-\xi)\ \mathbf{1}_{\{\varphi(X_{n+1})\geq\xi\}}|\mathcal{F}_n\right]_{|\xi=\xi_n} \\
&\qquad -\mathbb{E}\left[(\Psi(\varphi(X))-\xi)\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]^2_{|\xi=\xi_n}\Big) \\
&\xrightarrow{a.s.} \frac{1}{(1-\alpha)^2}\Big(\mathbb{E}\left[(\Psi(\varphi(X))-\xi^*)^2\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\
&\qquad -\mathbb{E}\left[(\Psi(\varphi(X))-\xi^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right]^2\Big) \\
&= \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X))-\xi^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right).
\end{aligned}$$

Using the continuity of both functions $\xi\mapsto\mathbb{E}\left[(\Psi(\varphi(X))-\xi)\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]$ and $\xi\mapsto\mathbb{E}\left[(\Psi(\varphi(X))-\xi)^2\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]$ at ξ^*, which follows from the continuity of Ψ and of the distribution function of $\varphi(X)$, finally yields the $a.s.$ convergence of $\mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right]$ toward

$$\Gamma = \begin{pmatrix} \frac{\alpha}{1-\alpha} & \frac{\alpha}{(1-\alpha)^2}\mathbb{E}\left[(\Psi(\varphi(X))-\xi^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\ \frac{\alpha}{(1-\alpha)^2}\mathbb{E}\left[(\Psi(\varphi(X))-\xi^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X))-\xi^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right) \end{pmatrix}.$$

If $\gamma_n = \frac{\gamma_1}{n^a}$ with $\gamma_1 > 0$ and $\frac{1}{2} < a < 1$, Ruppert-Polyak's Theorem implies that

$$\sqrt{n}\left(\bar{Z}_n - z^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma)$$

where $\Sigma = M^{-1}\Gamma\left(M^{-1}\right)^T$ is given by (2.19). This completes the proof. □

Remarks: • It is possible to replace $w(\xi,x)$ in (2.13) and (2.15) by $\tilde{w}(\xi,x) = \frac{1}{1-\alpha}\Psi(\varphi(x))\mathbf{1}_{\{\varphi(x)\geq\xi\}}$ since $C^* = \mathbb{E}\left[\tilde{w}\left(\xi^*,X\right)\right]$. Thus, we only have to change also

the martingale increment sequence $(\Delta N_n)_{n\geq 1}$ by $\left(\Delta\widetilde{N}_n\right)_{n\geq 1}$ defined by

$$\Delta\widetilde{N}_n := \frac{1}{1-\alpha}\left(\mathbb{E}\left[\Psi(\varphi(X))\mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]_{|\xi=\xi_{n-1}} - \Psi(\varphi(X_n))\mathbf{1}_{\{\varphi(X_n)\geq\xi_{n-1}\}}\right).$$

This provides another procedure $\tilde{C}_n$ for the computation of the Ψ-CVaR$_\alpha$ which satisfies a Gaussian CLT with the same asymptotic covariance matrix.

• The quantile estimate based on the inversion of the empirical distribution function satisfies a Gaussian CLT with the same asymptotic covariance matrix than the one of the procedure $\overline{\xi}_n$, see for example [84] p.75. Obviously, there is no reason to believe that this first version can do better than the empirical quantile estimate. However, our quantile estimate has the advantage to be recursive: it naturally combines with a recursive IS algorithm in an adaptive way. In terms of computational complexity, once N loss samples have been generated, the behaviour of the inversion of the empirical distribution function method needs a sorting algorithm: good behaviour is $\mathcal{O}(N\log(N))$ element comparisons to sort the list of loss samples. Whereas the behaviour of the recursive quantile algorithm is $\mathcal{O}(N)$. • One shows that if we choose $\beta_n = \frac{1}{n}$, $n \geq 1$ and $\gamma_n = \frac{1}{n^a}$ with $\frac{1}{2} < a < 1$ in (2.15), the resulting two-time scale procedure satisfies a Gaussian CLT with the same asymptotic covariance matrix Γ (at rates $\sqrt{\gamma_n^{-1}}$ and $\sqrt{n}$). However, by averaging the first component ξ_n, the resulting procedure becomes asymptotically efficient (*i.e.* rate $\sqrt{n}$).

Proposition 2.2.5. *(Estimation of variance and confidence interval) For every $n \geq 1$, set*

$$\sigma_n^2 := \frac{1}{(1-\alpha)^2}\left(\frac{1}{n}\sum_{k=1}^{n}\left(\Psi(\varphi(X_k)) - \xi_{k-1}\right)^2\mathbf{1}_{\{\varphi(X_k)\geq\xi_{k-1}\}} - \left(\frac{1}{n}\sum_{k=1}^{n}\left(\Psi(\varphi(X_k)) - \xi_{k-1}\right)\mathbf{1}_{\{\varphi(X_k)\geq\xi_{k-1}\}}\right)^2\right)$$

where $(\xi_n)_{n\geq 0}$ is the first component of (2.6). If $(A2)_a$ is satisfied for some $a \geq 2$, then

$$\sigma_n^2 \xrightarrow{a.s.} \frac{1}{(1-\alpha)^2}\ \mathrm{Var}\ \left(\left(\Psi(\varphi(X)) - \xi^*\right)\mathbf{1}_{\varphi(X)\geq\xi^*}\right)$$

and

$$\sqrt{n}\,\frac{C_n - C^*}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{2.22}$$

Proof. The proof follows from standard arguments already used in the proof of the *a.s.* convergence of the sequence $(C_n)_{n\geq 1}$ defined by (2.13). □

In practice, the convergence of the algorithm will be chaotic. The bottleneck of this algorithm is that it is only updated on rare events since it tries to measure the tail distribution of $\varphi(X)$: $\mathbb{P}(\varphi(X) > \mathrm{VaR}_\alpha) = 1 - \alpha \approx 0$. Another problem may be the simulation of $\varphi(X)$. In practice, we have to deal with large portfolios of complex derivative securities and options. Each evaluation may require a lot of computational efforts and takes a long time. So, for practical implementation it is necessary to combine the above procedure with variance reduction techniques to achieve accurate results at a reasonable cost. The most appropriate technique when dealing with rare events is IS.

2.2.3 Some background on IS using stochastic approximation algorithm

The second tool we want to introduce in this paper is a recursive IS procedure which increases the probability of simulations for which $\varphi(X)$ exceeds ξ. Our goal is to combine it adaptively with our first naive algorithm. Assume that X has an absolutely continuous distribution $\mathbb{P}_X(dx) = p(x)\lambda_d(dx)$ where λ_d denotes the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$. The main idea of importance sampling by translation applied to the computation of

$$\mathbb{E}[F(X)],$$

where $F \in L^2(\mathbb{P}_X)$ satisfies $\mathbb{P}(F(X) \neq 0) > 0$, is to use the invariance of the Lebesgue measure by translation, for every $\theta \in \mathbb{R}^d$,

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X+\theta)\frac{p(X+\theta)}{p(X)}\right], \tag{2.23}$$

and among all these random vectors with the same expectation, we want to select the one with the lowest variance, *i.e.* the one with lowest quadratic norm

$$Q(\theta) := \mathbb{E}\left[F^2(X+\theta)\frac{p^2(X+\theta)}{p^2(X)}\right] \leq +\infty, \quad \theta \in \mathbb{R}^d. \tag{2.24}$$

If the following assumption

$$\forall \theta \in \mathbb{R}^d, \qquad \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right] < +\infty \tag{B1}$$

holds true, then Q is everywhere finite and a reverse change of variable shows that:

$$Q(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right], \quad \theta \in \mathbb{R}^d. \tag{2.25}$$

Now if p satisfies

$$\begin{cases} (i) \ \ \forall x \in \mathbb{R}^d, \ \theta \mapsto p(x-\theta) \text{ is log-}\textit{concave} \\ (ii) \ \ \forall x \in \mathbb{R}^d, \ \lim_{|\theta|\to+\infty} p(x-\theta) = 0 \quad \text{or } \forall x \in \mathbb{R}^d, \ \lim_{|\theta|\to+\infty} \frac{p(x-\theta)}{p^2(x-\frac{\theta}{2})} = 0, \end{cases} \tag{B2}$$

one shows that Q is (strictly) finite, convex, goes to infinity at infinity so that $\arg\min Q = \{\nabla Q = 0\}$ is non empty (see [1] and [62]). Provided that ∇Q admits a representation as an expectation, then it is possible to devise a recursive RM procedure to approximate the optimal parameter θ^*. Recursive IS by stochastic approximation has been first investigated by Kushner and then by several authors, see e.g. [24] and [35] in order to "optimize" or "improve" the change of measure in IS using a stochastic gradient RM algorithm based on the representation of $\nabla Q(\theta)$. Recently, it has been brought back to light by Arouna (see [1]) in the Gaussian case, based on the natural representation of ∇Q obtained by formally differentiating (2.25). Since we have no knowledge about the regularity of F and do not wish

to have any, we differentiate the second representation of Q in (2.25) and not (2.24). We obtain $\nabla Q(\theta) = \mathbb{E}\left[K(\theta, X)\right]$.

When $X = \mathcal{N}(0,1)$, $Q(\theta) = e^{\frac{|\theta|^2}{2}}\mathbb{E}[F^2(X)e^{-\theta X}]$ so that $K(\theta, x) = e^{\frac{|\theta|^2}{2}}F^2(x)e^{-\theta x}(\theta - x)$. However, given this resulting form of K, the classical convergence results do not apply since $||K(\theta, X)||_2$ is not sub-linear in θ (see condition (2.8) of Theorem 2.2). This induces the explosion of the procedure at almost every implementation as pointed out in [1]. This leads the author to introduce a "constrained" variant of the regular procedure based on repeated reinitializations known as the projection "à la Chen". It forces the stability of the algorithm and prevents explosion. Let us also mention a first alternative approach investigated in [1] and [2], where Arouna and Bardou change the function to be minimized by introducing an entropy based criterion. Although it is only an approximation, it turns out to be often close to the original method.

Recently, Lemaire and Pagès in [62] revisited the original approach and provided a new representation of $\nabla Q(\theta)$ for which the resulting $K(\theta, X)$ has a linear growth in θ so that all assumptions of Theorem 2.2 are satisfied. Thanks to *a third translation* of the variable θ, it is possible to plug back the parameter θ "into" F, the function F having in common applications a known behaviour at infinity which makes possible to devise a "regular" and "unconstrained" stochastic algorithm. We will rely partially on this approach to devise our final procedure to compute both VaR and CVaR. To be more specific about the methodology proposed in [62], we introduce the following assumption on the probability density p of X

$$\exists b \in [1,2] \text{ such that } \begin{cases} (i) \quad \frac{|\nabla p(x)|}{p(x)} = O(|x|^{b-1}) \quad \text{as} \quad |x| \to \infty \\ \\ (ii) \quad \exists \rho > 0, \log\left(p(x)\right) + \rho|x|^b \text{ is convex,} \end{cases} \tag{B3}$$

and introduce the assumption on F :

$$\forall A > 0, \mathbb{E}\left[F(X)^2 e^{A|X|^{b-1}}\right] < +\infty. \tag{B4}$$

One shows that as soon as (B1), (B2), (B3) and (B4) are satisfied, Q_1 and Q_2 are both finite and differentiable on $\mathbb{R}^d$ with a gradient given by

$$\nabla Q(\theta) \quad := \quad \mathbb{E}\left[F(X-\theta)^2 \underbrace{\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}}_{W(\theta,X)}\right]. \tag{2.26}$$

This expression may look complicated at first glance but in fact the weight term $W(\theta, X)$ can be easily controlled by a deterministic function of θ since

$$|W(\theta, X)| \leq e^{2\rho|\theta|^b}(A|x|^{b-1} + A|\theta|^{b-1} + B) \tag{2.27}$$

for some real constants A and B. In the case of a normal distribution $X \overset{d}{=} \mathcal{N}(0;1)$,

$$W(\theta, X) = e^{\theta^2}(2\theta - X).$$

So, if we have a control on the growth of the function F, typically for some positive constant c

$$\begin{cases} \forall x \in \mathbb{R}^d, |F(x)| \leq G(x) \quad \text{and} \quad G(x+y) \leq C(1+G(x))^c(1+G(y))^c \\ \\ \mathbb{E}\left[|X|^{2(b-1)}G(X)^{4c}\right] < +\infty, \end{cases} \tag{B5}$$

then by setting

$$\widetilde{W}(\theta, X) := \frac{e^{-2\rho|\theta|^b}}{1+G(-\theta)^{2c}} W(\theta, X), \tag{2.28}$$

we can define K by

$$K(\theta, x) := F(x-\theta)^2 \widetilde{W}(\theta, X) \tag{2.29}$$

so that it satisfies the linear growth assumption (2.8) of Theorem 2.2 and

$$\left\{\theta \in \mathbb{R}^d \mid \mathbb{E}\left[K(\theta, X)\right] = 0\right\} = \left\{\theta \in \mathbb{R}^d \mid \nabla Q(\theta) = 0\right\}.$$

Moreover, since Q is convex ∇Q satisfies (2.7). Now we are in position to derive a recursive unconstrained RM algorithm

$$\theta_n = \theta_{n-1} - \gamma_n K(\theta_{n-1}, X_n), \ \theta_0 \in \mathbb{R}^d, \tag{2.30}$$

that *a.s.* converges to an $\arg\min Q$-valued (square integrable) random variable θ^*.

2.3 Design of a faster procedure: importance sampling and moving confidence level

2.3.1 Unconstrained adaptive importance sampling device

We noted previously that the bottleneck in using the above algorithm lies in its very slow and chaotic convergence owing to the fact that $\mathbb{P}(\varphi(X) > \xi^*) = 1-\alpha$ is close to 0. This means that we observe fewer and fewer simulations for which $\varphi(X_k) > \xi_{k-1}$ as the algorithm evolves. Thus, it becomes more and more difficult to compute efficiently some estimates of VaR_α and CVaR_α when $\alpha \approx 1$. Moreover, in the bank and energy sectors, practitioners usually deal with huge portfolio made of hundreds or thousands of risk factors and options. The evaluation step of $\varphi(X)$ may be extremely time consuming. Consequently, to achieve accurate estimates of both VaR_α and CVaR_α with reasonable computational effort, the above algorithm (2.16) drastically needs to be speeded up by an IS procedure to "recenter" the simulations where "things do happen", *i.e.* which generates scenarios for which $\varphi(X)$ exceeds ξ.

In this section we will focus on IS by mean translation. Our aim is to combine adaptively the IS (unconstrained) recursive procedure investigated in [62] with our first "naive" approach described in (2.16). Doing so every new sample is used to both optimize the IS change of measure and update VaR and CVaR procedures. We plan to minimize the asymptotic variance of both components of the algorithm (in its "averaged" form, as detailed in Theorem 2.4), namely

$$\frac{\alpha(1-\alpha)}{f^2_{\varphi(X)}(\xi^*)} = \frac{\mathrm{Var}(\mathbf{1}_{\varphi(X)\geq\xi^*})}{f^2_{\varphi(X)}(\xi^*)} \quad \text{for the } \mathrm{VaR}_\alpha, \tag{2.31}$$

and,

$$\frac{\mathrm{Var}((\Psi(\varphi(X)) - \xi^*)\mathbf{1}_{\varphi(\mathbf{X})\geq\xi^*})}{(1-\alpha)^2} \quad \text{for the } \mathrm{CVaR}_\alpha, \tag{2.32}$$

provided the non-degeneracy assumption

$$\forall \xi \in \arg\min V,\ \mathbb{P}\left(\left(\Psi(\varphi(X)) - \xi\right)^2 \mathbf{1}_{\{\varphi(X)\geq\xi\}} > 0\right) > 0, \tag{A3}$$

holds. Since the density $f_{\varphi(X)}(\xi^*)$ is an intrinsic constant (and comes in fact from the Jacobian matrix $Dh(\xi^*, C^*)$ of the mean function h of the algorithm) we are led to apply the IS paradigm described in Section 2.3 to

$$F_1^*(X) = \mathbf{1}_{\varphi(X)\geq\xi^*} \quad \text{and} \quad F_2^*(X) = (\Psi(\varphi(X)) - \xi^*)\,\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}.$$

Let us temporary forget that of course we do not know ξ^* at this stage. Those two functionals are related to the minimization of the two convex functions

$$Q_1(\theta, \xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p(X-\theta)}\right] \tag{2.33}$$

$$Q_2(\mu, \xi^*) := \mathbb{E}\left[\left(\Psi(\varphi(X)) - \xi^*\right)^2 \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p(X-\mu)}\right]. \tag{2.34}$$

We can apply to these functions the minimizing procedure (2.30) described at section 2.3. Since

$$H_1(\xi^*, x) = 1 - \frac{1}{1-\alpha}F_1^*(x) \quad \text{and} \quad H_2(\xi^*, C^*, x) = C^* - \Psi(\xi^*) - \frac{1}{1-\alpha}F_2^*(x) \tag{2.35}$$

it is clear, owing to (2.23) that

$$\mathbb{E}\left[H_i(\xi^*, X)\right] = \mathbb{E}\left[H_i(\xi^*, X+\theta)\frac{p(X+\theta)}{p(X)}\right] \quad i = 1,\ 2.$$

Now, since we do not know either ξ^* and C^* (the VaR_α and the CVaR_α) respectively we make the whole procedure adaptive by replacing at step n, these unknown parameters by their running approximation at step $n-1$. This finally justifies to introduce the following global procedure. One defines the state variable, for $n \geq 0$,

$$Z_n := (\xi_n, C_n, \theta_n, \mu_n),$$

where ξ_n, C_n denotes the VaR_α and the CVaR_α approximate, θ_n, μ_n denotes the variance reducers for the VaR and the CVaR procedures. We update this state variable recursively by

$$Z_n = Z_{n-1} - \gamma_n L(Z_{n-1}, X_n), \tag{2.36}$$

where $(X_n)_{n\geq1}$ is an i.i.d. sequence with distributions X (and probability density

p) and

$$\begin{aligned}
L_1(\xi,\theta,x) &:= e^{-\rho|\theta|^b}\left(1-\frac{1}{1-\alpha}\,\mathbf{1}_{\{\varphi(x+\theta)\geq\xi\}}\frac{p(x+\theta)}{p(x)}\right),\\
L_2(\xi,C,\mu,x) &:= C-\Psi(\xi)-\frac{1}{1-\alpha}(\Psi(\varphi(x+\mu))-\xi)\,\mathbf{1}_{\{\varphi(x+\mu)\geq\xi\}}\frac{p(x+\mu)}{p(x)},\\
L_3(\xi,\theta,x) &:= e^{-2\rho|\theta|^b}\mathbf{1}_{\{\varphi(x-\theta)\geq\xi\}}\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\frac{\nabla p(x-2\theta)}{p(x-2\theta)},\\
L_4(\xi,\mu,x) &:= \frac{e^{-2\rho|\mu|^b}}{1+G(-\mu)^{2c}+\Psi(\xi)^2}\left(\Psi(\varphi(x-\mu))-\xi\right)^2\\
&\qquad\times\mathbf{1}_{\{\varphi(x-\mu)\geq\xi\}}\frac{p^2(x-\mu)}{p(x)p(x-2\mu)}\frac{\nabla p(x-2\mu)}{p(x-2\mu)}.
\end{aligned}\tag{2.37}$$

The following proposition establishes the *a.s.* convergence of the procedure. For the sake of simplicity we will assume the uniqueness of the VaR_α of $\varphi(X)$.

Proposition 2.3.1. *(Efficient computation of VaR and CVaR). Suppose that $\Psi(\varphi(X))\in L^2(\mathbb{P})$, that the distribution function of $\varphi(X)$ is continuous and increasing (so that $VaR_\alpha(\varphi(X))$ is unique) and that (A3) holds. Assume that, for every $\xi\in\mathbb{R}$, $Q_i(.,\xi)$ (i=1,2) satisfies (B1), i.e.*

$$\forall\theta\in\mathbb{R}^d,\quad \mathbb{E}\left[\left(1+\left(\Psi(\varphi(X))-\xi\right)^2\right)\boldsymbol{1}_{\varphi(X)\geq\xi}\frac{p(X)}{p(X-\theta)}\right]<+\infty.\tag{2.38}$$

Suppose that p satisfies (B2) and (B3) and that

$$\forall A>0,\mathbb{E}\left[\left(\Psi(\varphi(X))^2+1\right)e^{A|X|^{b-1}}\right]<+\infty.$$

Assume that the step sequence $(\gamma_n)_{n\geq1}$ satisfies (A1). Then,

$$Z_n\xrightarrow{a.s.}z^*:=(\xi^*,C^*,\theta^*_\alpha,\mu^*_\alpha)$$

where $\xi^*=\mathrm{VaR}_\alpha(\varphi(X))$, $C^*=\Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X))$ and $(\theta^*_\alpha,\mu^*_\alpha)$ are the optimal variance reducers (to be precise some random vectors taking values in $\{\nabla Q_1(\xi^*,.)=0\}$ and $\{\nabla Q_2(\xi^*,.)=0\}$ respectively).

Proof. We first prove the *a.s.* convergence of the 3-tuple (ξ_n,θ_n,μ_n) that of $(C_n)_{n\geq1}$ will follow by the same arguments used in the proof in Section 2.2. The mean function l is defined by

$$l(\xi,\theta,\mu):=\left(1-\frac{1}{1-\alpha}\mathbb{P}\left(\varphi(X)\geq\xi\right),e^{-2\rho|\theta|^b}\nabla Q_1\left(\theta,\xi\right),\frac{e^{-2\rho|\mu|^b}}{1+G(-\mu)^c+\Psi(\xi)^2}\nabla Q_2(\mu,\xi)\right),$$

hence,

$$\mathcal{T}^*=\{l=0\}=\{\xi^*\}\times\{\nabla Q_1(\xi^*,.)=0\}\times\{\nabla Q_2(\xi^*,.)=0\}.$$

In order to apply the extended Robbins-Monro Theorem, we have to check the following facts:

• *Mean reversion*: One checks that $\forall\zeta=(\xi,\theta,\mu)\in\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d\setminus\mathcal{T}^*$, $\forall\zeta^*\in\mathcal{T}^*$,

$$\begin{aligned}\langle \zeta - \zeta^*, l(\zeta)\rangle &= (\xi-\xi^*)\frac{(\mathbb{P}(\varphi(X)\le\xi)-\alpha)}{1-\alpha} + \frac{e^{-2\rho|\theta|^b}}{1-\alpha}\langle \theta-\theta^*_\alpha, \nabla Q_1(\theta,\xi)\rangle \\ &\quad + \frac{e^{-2\rho|\mu|^b}}{(1-\alpha)(1+F(-\mu)^{2c})}\langle \mu-\mu^*_\alpha, \nabla Q_2(\mu,\xi)\rangle > 0,\end{aligned}$$

owing to the convexity of $\theta \mapsto Q_1(\theta,\xi)$ and $\mu \mapsto Q_2(\mu,\xi)$, for every $\xi \in \mathbb{R}$.

• *Linear growth:* Let us first deal with L_1. First note that:

$$\begin{aligned}\mathbb{E}\left[L_1(\xi,\theta,X)^2\right] &\le C\left(1+\mathbb{E}\left[e^{-2\rho|\theta|^b}\mathbf{1}_{\{\varphi(X+\theta)\ge\xi\}}\frac{p^2(X+\theta)}{p^2(X)}\right]\right) \\ &\le C\left(1+\mathbb{E}\left[e^{-2\rho|\theta|^b}\frac{p(X)}{p(X-\theta)}\right]\right).\end{aligned}$$

Now, elementary computations show (see [62] for more details) that (B3)(ii) implies that

$$\frac{p^2(x)}{p(x-\theta)} \le e^{2\rho|\theta|^b}p(x+\theta),$$

so that

$$\mathbb{E}\left[e^{-2\rho|\theta|^b}\frac{p(X)}{p(X-\theta)}\right] \le \mathbb{E}\left[\frac{p(X+\theta)}{p(X)}\right] = 1.$$

L_3 and L_4 can be treated by a straightforward adaptation of the proofs in [62]. Then, one can apply Theorem 2.2 which yields the announced result for (ξ_n,θ_n,μ_n). The *a.s.* convergence of C_n toward C^* can be deduced from the *a.s.* convergence of the series

$$M_n^\gamma := \sum_{k=1}^n \gamma_k \Delta\widetilde{M}_k,\ n \ge 1,$$

where $\Delta\widetilde{M}_n$ are martingale increments defined by

$$\begin{aligned}\Delta\widetilde{M}_n &= \mathbb{E}[(\Psi(\varphi(X))-\xi)\,\mathbf{1}_{\{\varphi(X)\ge\xi\}}]_{|\xi=\xi_{n-1}} \\ &\quad -(\Psi(\varphi(X_n+\mu_{n-1}))-\xi_{n-1})\,\mathbf{1}_{\{\varphi(X_n+\mu_{n-1})\ge\xi_{n-1}\}}\frac{p(X_n+\mu_{n-1})}{p(X_n)},\quad n\ge 1,\end{aligned}$$

satisfying

$$\mathbb{E}\left[\Delta\widetilde{M}_n^2|\mathcal{F}_{n-1}\right] \le \mathbb{E}\left[(\Psi(\varphi(X+\mu))-\xi)\,\mathbf{1}_{\{\varphi(X+\mu)\ge\xi\}}\frac{p(X+\mu)}{p(X)}\right]_{|\xi=\xi_{n-1},\theta=\theta_{n-1},\mu=\mu_{n-1}}.$$

We conclude by the same arguments used in the proof in Section 2.2. □

Now, we are interested by the rate of convergence of the procedure. It shows that the algorithm behaves as expected under quite standard assumptions: it satisfies a Gaussian CLT with optimal rate and minimal variances.

Theorem 2.3.2. *Suppose the assumptions of Proposition 3.1 hold true. Assume that $\Psi(\varphi(X)) \in L^{2a}(\mathbb{P})$ for some $a > 1$ and that the step sequence is $\gamma_n = \frac{\gamma_1}{n^p}$ with $\frac{1}{2} < p < 1$ and $\gamma_1 > 0$. Suppose furthermore that the density $f_{\varphi(X)}$ is differentiable*

and strictly positive on its support. Let $(\overline{\xi}_n, \overline{C}_n)_{n\geq 1}$ *be the sequence of Cesaro means defined by:*

$$\overline{\xi}_n := \frac{\xi_0 + ... + \xi_{n-1}}{n}, \quad \overline{C}_n := \frac{C_0 + ... + C_{n-1}}{n}, \quad n \geq 1.$$

This sequence satisfies the following CLT:

$$\sqrt{n}\begin{pmatrix} \overline{\xi}_n - \xi^* \\ \overline{C}_n - C^* \end{pmatrix} \overset{\mathcal{L}}{\to} \mathcal{N}(0, \Sigma^*) \qquad as\ n \to +\infty, \tag{2.39}$$

where

$$\begin{aligned}
\Sigma^*_{1,1} &= \frac{1}{f^2_{\varphi(X)}(\xi^*)}\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)\geq\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right), \\
\Sigma^*_{1,2} &= \Sigma^*_{2,1} = \frac{1}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathrm{Cov}\left((\Psi(\varphi(X+\mu^*_\alpha)) - \xi^*)\,\mathbf{1}_{\{\varphi(X+\mu^*_\alpha)>\xi^*\}}\frac{p(X+\mu^*_\alpha)}{p(X)},\right. \\
&\qquad \left.\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)\geq\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right), \\
\Sigma^*_{2,2} &= \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X+\mu^*_\alpha)) - \xi^*)\ \mathbf{1}_{\{\varphi(X+\mu^*_\alpha)\geq\xi^*\}}\frac{p(X+\mu^*_\alpha)}{p(X)}\right).
\end{aligned}$$

Proof. The proof is built like the one of Theorem 2.4. If we denote h the mean function of the global algorithm $h(z) = \mathbb{E}[L(z, X)]$, the algorithm (2.36) can be written as

$$Z_n = Z_{n-1} - \gamma_n\left(h(Z_{n-1}) + \tilde{\epsilon}_n\right),\ n \geq 1, \quad Z_0 = (\xi_0, 0)\,,\ \xi_0 \in L^1(\mathbb{P}), \tag{2.40}$$

where the first two components of h are the same function as the ones in the proof of Theorem 2.4 and $(\tilde{\epsilon}_n)_{n\geq 1}$ denotes the $\mathcal{F}_n$-adapted martingale increments sequence where

$$\begin{aligned}
\tilde{\epsilon}_{1,n} &:= \frac{1}{1-\alpha}\left(\mathbb{P}\left(\varphi(X) \geq \xi\right)_{|\xi=\xi_n} - \ \mathbf{1}_{\{\varphi(X_{n+1}+\theta_n)\geq\xi_n\}}\frac{p(X_{n+1}+\theta_n)}{p(X_{n+1})}\right), \\
\tilde{\epsilon}_{2,n} &:= \frac{1}{1-\alpha}\left(\mathbb{E}[(\Psi(\varphi(X)) - \xi)\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}]_{|\xi=\xi_n}\right. \\
&\qquad \left. -(\Psi(\varphi(X_{n+1}+\mu_n)) - \xi_n)\ \mathbf{1}_{\{\varphi(X_{n+1}+\mu_n)\geq\xi_n\}}\frac{p(X_{n+1}+\mu_n)}{p(X_{n+1})}\right).
\end{aligned}$$

One can check easily that the sequence $(\tilde{\epsilon}_n)_{n\geq 1}$ satisfies $(i) - (iv)$ of (2.17). □

Remarks: • There exists a CLT for the whole sequence $(Z_n)_{n\geq 1}$ and for its empirical mean $(\overline{Z}_n)_{n\geq 1}$ according to Ruppert and Polyak averaging principle. We only stated the result for the two components of interest (the ones which converge to VaR and CVaR respectively) since we only need rough estimates for the other two (see below).

• In the first Central Limit Theorem (Theorem 2.4) for quantile estimation, the factor $\alpha(1-\alpha)$ is the variance of the indicator function of the event $\{\varphi(X) \geq \xi^*\}$. With our recursive IS procedure, it is replaced by the variance of the shifted indicator function modified by the measure change: $\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)>\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right)$. For

further details on the rate of convergence of the unconstrained recursive importance sampling procedure, we refer to [62].

Now, let us point out an important issue. The algorithm (2.36) raises an important problem numerically speaking. Actually, we have two algorithm ξ_n and (θ_n, μ_n) that are in competitive conditions, *i.e.* on one hand, we added an IS procedure to $(\xi_n)_{n\geq 1}$ to improve the convergence toward ξ^*, and on the other hand, the adjustment of the parameters (θ_n, μ_n) "need" some samples X_{n+1} satisfying $\varphi(X_{n+1} - \theta_n) > \xi_n$ and $\varphi(X_{n+1} - \mu_n) > \xi_n$ ($\Psi \equiv Id$) which tend to become rare events. Somehow, we postponed the problems resulting from rare events on the IS procedure itself which may "freeze". This in term suggests to break the link between the VaR-CVaR and the IS procedures by introducing a VaR *companion procedure* that will drive the IS parameters to the tail distribution. A solution to do this is to make the confidence level increase slowly from a lower value (say $\alpha_0 = 50\%$) up to the target level α. This kind of incremental threshold increase has been already proposed in [55] in a different framework. This idea is developed in the next section.

2.3.2 How to control the move towards the critical risk area: the final procedure

From a theoretical point of view, so far, we considered the purely adaptive approach where we approximate $(\xi^*, C^*, \theta^*_\alpha, \mu^*_\alpha)$ using the *same innovation sequences*. From a numerical point of view, we only need a rough estimate of the optimal IS parameters $(\theta^*_\alpha, \mu^*_\alpha)$. So that we are led to break the algorithm into two phases. Firstly, we compute a rough estimate of the optimal IS parameters (θ_M, μ_M) with a small number of iterations M and in a second time, estimate the VaR_α and the CVaR_α with those optimized parameters with N iterations ($M \ll N$ in practice).

Now, in order to circumvent the problem induced by the IS procedure, we propose to introduce companion VaR procedure (without IS, *i.e.*, based on H_1 from Section 2.2) that will lead the IS parameters into the critical risk area during a first phase of the simulation, say the first M iterations. An idea to control the growth of θ_n and μ_n at the beginning of the algorithm, since we have no idea on how to twist the distribution of $\varphi(X)$, is to move slowly toward the target critical risk area (at level α) in which $\varphi(X)$ exceeds ξ by introducing a non-decreasing sequence α_n slowly converging to α during the first phase. Since the algorithm for the CVaR component C_n is free of α, by doing so, we only modify the VaR procedure ξ_n. The function H_1 in (2.16) is replaced by its counterpart which depends on the moving confidence level α_n, namely

$$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n \hat{H}_1\left(\hat{\xi}_{n-1}, X_n, \alpha_n\right), \quad n \geq 1, \hat{\xi}_0 = \xi_0 \in L^1(\mathbb{P}). \tag{2.41}$$

where,

$$\forall\, \xi \in \mathbb{R},\ \forall\, x \in \mathbb{R}^d, \forall\, \hat{\alpha} \in]0,1[,\ \hat{H}_1(\xi, x, \hat{\alpha}) = 1 - \frac{1}{1-\hat{\alpha}}\, \mathbf{1}_{\{\varphi(x)\geq \xi\}}.$$

The sequence $\left(\hat{\xi}_n\right)_{n\geq 0}$ is only designed to drive "smoothly" the IS procedures toward the "critical area" at the beginning of the procedure, say during the first M iterations

and in no case to approximate ξ^* or C^*. To be more precise, we define recursively the variance reducer sequence $(\hat{\theta}_n)_{n\geq 1}$, $(\hat{\mu}_n)_{n\geq 1}$ by plugging at each step n, $\hat{\xi}_{n-1}$ into $L_3(.,\hat{\theta}_{n-1},X_n)$ and $L_4(.,\hat{\mu}_{n-1},X_n)$ as defined in Section 3.1. This reads as follows, for $n \geq 1$,

$$\begin{cases} \hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n \hat{H}_1\left(\hat{\xi}_{n-1}, X_n, \alpha_n\right), \quad \hat{\xi}_0 \in L^1(\mathbb{P}), \\ \hat{\theta}_n = \hat{\theta}_{n-1} - \gamma_n L_3\left(\hat{\xi}_{n-1}, \hat{\theta}_{n-1}, X_n\right), \quad \theta_0 \in \mathbb{R}^d, \\ \hat{\mu}_n = \hat{\mu}_{n-1} - \gamma_n L_4\left(\hat{\xi}_{n-1}, \hat{\mu}_{n-1}, X_n\right), \quad \mu_0 \in \mathbb{R}^d. \end{cases} \tag{2.42}$$

Although, we are not really interested in the asymptotic of this procedure $(\hat{\xi}_n)$, its theoretical convergence follows from Theorem 2.2: as a matter of fact if we define a remainder term r_n by:

$$r_n := \hat{H}_1\left(\hat{\xi}_{n-1}, X_n, \alpha_n\right) - H_1\left(\hat{\xi}_{n-1}, X_n\right), \quad n \geq 1,$$

the procedure defined by (2.42) now reads

$$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n(H_1(\hat{\xi}_{n-1}, X_n) + r_n), \quad n \geq 1, \quad \hat{\xi}_0 \in L^1(\mathbb{P}). \tag{2.43}$$

One checks that

$$|r_n| \leq \frac{|\alpha_n - \alpha|}{(1-\alpha)^2},$$

so that Assumption (2.9) of Theorem 2.2 is satisfied as soon as

$$\sum_{n\geq 1} \gamma_n(\alpha - \alpha_n)^2 < +\infty.$$

2.3.3 A final procedure for practical implementation

In practice, we divide our procedure into two phases:
▷ Phase I is devoted to the estimation of the variance reducers $(\theta^*_\alpha, \mu^*_\alpha)$ using (2.42). The moving confidence level α has been settled as follows ($M \approx 15000$) :

- $\alpha_n = 50\%$ for $1 \leq n \leq M_1 := M/3$,
- $\alpha_n = 80\%$ for $M_1 < n \leq 2M_1$,
- and $\alpha_n = \alpha$ for $2M_1 < n \leq M$.

▷ Phase II produces some estimates for (ξ^*, C^*) based on the procedure defined by (2.36) and its Cesaro mean with N iterations. Note that during this phase, we keep on updating the IS parameters adaptively.

Now, we can summarize the two phase of the final procedure by the following pseudo-code:

An alternative, especially as concerns practical implementation, is to replace to Phase II by

Phase II' in which the variance reducers coming from Phase I are frozen at $\hat{\theta}_M$, $\hat{\mu}_M$. The only updated sequence is (ξ_n, C_n), as follows

$$\begin{aligned} \xi_n &= \xi_{n-1} - \gamma_n L_1\left(\xi_{n-1}, \hat{\theta}_M, X_n\right), \\ C_n &= C_{n-1} - \gamma_n L_2\left(\xi_{n-1}, C_{n-1}, \hat{\mu}_M, X_n\right). \end{aligned}$$

Phase I: Estimation of $(\mu_\alpha^*, \theta_\alpha^*)$. $M \ll N$ (typically $M \approx N/100$).
for $n = 1$ to M **do**
$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n \hat{H}_1\left(\hat{\xi}_{n-1}, X_n, \alpha_n\right)$,
$\hat{\theta}_n = \hat{\theta}_{n-1} - \gamma_n L_3\left(\hat{\xi}_{n-1}, \hat{\theta}_{n-1}, X_n\right)$,
$\hat{\mu}_n = \hat{\mu}_{n-1} - \gamma_n L_4\left(\hat{\xi}_{n-1}, \hat{\mu}_{n-1}, X_n\right)$.
end for

Phase II: Estimation of (ξ^*, C^*). Set, for instance, $\xi_0 = \hat{\xi}_M$, $C_0 = 0$, $\theta_0 = \hat{\theta}_M$, and $\mu_0 = \hat{\mu}_M$.
for $n = 1$ to N **do**
$\xi_n = \xi_{n-1} - \gamma_n L_1\left(\xi_{n-1}, \theta_{n-1}, X_n\right)$,
$C_n = C_{n-1} - \gamma_n L_2\left(\xi_{n-1}, C_{n-1}, \mu_{n-1}, X_n\right)$,
$\theta_n = \theta_{n-1} - \gamma_n L_3\left(\xi_{n-1}, \theta_{n-1}, X_n\right)$,
$\mu_n = \mu_{n-1} - \gamma_n L_4\left(\xi_{n-1}, \mu_{n-1}, X_n\right)$,

Compute the Cesaro means

$\bar{\xi}_n = \bar{\xi}_{n-1} - \frac{1}{n}\left(\bar{\xi}_{n-1} - \xi_n\right)$,
$\bar{C}_n = \bar{C}_{n-1} - \frac{1}{n}\left(\bar{C}_{n-1} - C_n\right)$.
end for

(ξ^*, C^*) is estimated by $(\bar{\xi}_N, \bar{C}_N)$.

2.4 Towards some extensions

2.4.1 Extension to exponential change of measure: the Esscher transform

Considering an exponential change of measure (also called Esscher transform) instead of the mean translation is a rather natural idea that has already been investigated in [50] and [62] to extend the constrained IS stochastic approximation algorithm with repeated projections introduced in [1]. We briefly introduce the framework and give the main results without any proofs (for more details, see [62]). Let ψ denote the cumulant generating function (or log-Laplace) of X *i.e.* the function defined by $\psi(\theta) := \log \mathbb{E}[e^{\langle \theta, X\rangle}]$. We assume that $\psi(\theta) < +\infty$, which implies that ψ is an infinitely differentiable convex function and define

$$p_\theta(x) = e^{\langle \theta, x\rangle - \psi(\theta)} p(x), \quad x \in \mathbb{R}^d.$$

We denote by $X^{(\theta)}$ any random variable with distribution p_θ. We make the following assumption on the function ψ

$$\lim_{|\theta|} \psi(\theta) - 2\psi\left(\frac{\theta}{2}\right) = +\infty \quad \text{and} \quad \exists \delta > 0,\ \theta \mapsto \psi(\theta) - \delta|\theta|^2 \text{ is concave.} \qquad (H_\delta^{es})$$

The two functionals to be minimized are

$$Q_1(\theta, \xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X) > \xi^*\}} e^{-\langle \theta, X\rangle + \psi(\theta)}\right] \qquad (2.44)$$

$$Q_2(\mu, \xi^*) := \mathbb{E}\left[(\Psi(\varphi(X)) - \xi^*)^2 \mathbf{1}_{\{\varphi(X) > \xi^*\}} e^{-\langle \mu, X\rangle + \psi(\mu)}\right]. \qquad (2.45)$$

According to Proposition 3 in [62] as soon as ψ satisfies (H_δ^{es}) and that,

$$\forall \xi \in \mathbb{R}, \forall \theta \in \mathbb{R}^d, \quad \mathbb{E}[|X| \left(1 + \Psi(\varphi(X))^2\right) e^{\langle \theta, X \rangle}] < +\infty, \tag{2.46}$$

for every $\xi \in \mathbb{R}$, the functions $Q_1(., \xi)$ and $Q_2(., \xi)$ are finite, convex, differentiable on $\mathbb{R}^d$, go to infinity at infinity, so that $\arg\min Q_1(., \xi)$ and $\arg\min Q_2(., \xi)$ are non empty. Moreover, their gradients are given by

$$\nabla_\theta Q_1(\theta, \xi) = \mathbb{E}\left[(\nabla\psi(\theta) - X^{(-\theta)})\mathbf{1}_{\left\{\varphi(X^{(-\theta)})>\xi\right\}}\right] e^{\psi(\theta)-\psi(-\theta)} \tag{2.47}$$

$$\nabla_\mu Q_2(\mu, \xi) = \mathbb{E}\left[(\nabla\psi(\mu) - X^{(-\mu)})(\Psi(\varphi(X^{(-\mu)})) - \xi)^2\mathbf{1}_{\left\{\varphi(X^{(-\mu)})>\xi\right\}}\right] e^{\psi(\mu)-\psi(-\mu)} \tag{2.48}$$

with $\nabla\psi(\theta) = \frac{\mathbb{E}[Xe^{\langle\theta,X\rangle}]}{\mathbb{E}[e^{\langle\theta,X\rangle}]}$. Now, the main result of this section is the following theorem (for more details, we refer to [62]).

Theorem 2.4.1. *Suppose that ψ satisfies (H_δ^{es}) and that $(A2)_1$, (A3) hold. Assume that (2.46) is fulfilled and that*

$$\forall x \in \mathbb{R}^d, |\Psi(\varphi(x))| \leq Ce^{\frac{\lambda}{4}|x|} \quad \text{and} \quad \mathbb{E}[|X|^2 e^{\lambda|X|}] < +\infty.$$

One considers the recursive procedure

$$Z_n = Z_{n-1} - \gamma_n L(Z_{n-1}, X_n),\ n \geq 1,\ Z_0 = (\xi_0, C_0, \theta_0, \mu_0) \tag{2.49}$$

where $(\gamma_n)_{n\geq 1}$ satisfies the usual step assumption (A1), $Z_n := (\xi_n, C_n, \theta_n, \mu_n)$ and each component of L is defined by

$$L_1\left(\xi_{n-1}, \theta_{n-1}, X_n^{(\theta_{n-1})}\right) := e^{-\frac{\psi(\theta_{n-1})+\psi(-\theta_{n-1})}{2}}\left(1 - \frac{1}{1-\alpha}\mathbf{1}_{\left\{\varphi(X_n^{(\theta_{n-1})})>\xi_{n-1}\right\}} \times e^{\psi(\theta_{n-1})-\left\langle X_n^{(\theta_{n-1})}, \theta_{n-1}\right\rangle}\right),$$

$$L_2\left(\xi_{n-1}, C_{n-1}, \mu_{n-1}, X_n^{(\mu_{n-1})}\right) := C - \bar{w}(\xi_{n-1}, \mu_{n-1}, X_n^{(\mu_{n-1})}),$$

$$L_3\left(\xi_{n-1}, \theta_{n-1}, X^{(-\theta_{n-1})}\right) := \mathbf{1}_{\left\{\varphi(X^{(-\theta_{n-1})})>\xi_{n-1}\right\}}\left(\nabla\psi(\theta_{n-1}) - X^{(-\theta_{n-1})}\right),$$

$$L_4\left(\xi_{n-1}, \mu_{n-1}, X^{(-\mu_{n-1})}\right) := \frac{e^{-\frac{\lambda}{2}\sqrt{d}|\nabla\psi(-\mu_{n-1})|}}{1 + \Psi(\xi_{n-1})^2}(\Psi(\varphi(X^{(-\mu_{n-1})})) - \xi_{n-1})^2 \times \mathbf{1}_{\left\{\varphi(X^{(-\mu_{n-1})})>\xi_{n-1}\right\}}(\nabla\psi(\mu_{n-1}) - X^{(-\mu_{n-1})}),$$

with $\bar{w}(\xi, \mu, x) := \xi + \frac{1}{1-\alpha}(\Psi(\varphi(x)) - \xi)\mathbf{1}_{\{\varphi(x)>\xi\}}\ e^{\psi(\mu)-\langle\mu,x\rangle}$.

Then, Z_n converges a.s. toward $z^* := (\xi^*, C^*, \theta_\alpha^*, \mu_\alpha^*)$, where $\xi^* = \mathrm{VaR}_\alpha(\varphi(X))$, $C^* = \Psi\text{-CVaR}_\alpha(\varphi(X))$, θ_α^* is a (square integrable) $\arg\min Q_1(., \xi^*)$-valued random vector and μ_α^* is a (square integrable) $\arg\min Q_2(., \xi^*)$-valued random vector.

2.4.2 Extension to infinite dimensional setting

In the above sections, we proposed our algorithm in a finite dimensional setting where the value of the loss $L = \varphi(X)$ is a function of a random vector having values in $\mathbb{R}^d$. This is due to the fact that generally the value of a portfolio may

depend on a finite number of decisions taken in the past. Thus, the value of the loss at the horizon time $T-t$ may depend on a large number of dates in the past $t_0 = t < t_1 < t_2, ... < t_N = T - t$, with $N = 250$ for a portfolio with time interval $T - t = 1$ year. For instance, if we consider a simple portfolio composed of short positions on 250 calls with a maturity at each t_k and a strike K. The loss at time $t_N = 1$ year can be written:

$$L = \sum_{k=1}^{N} e^{r(t_N - t_k)}(S_{t_k} - K)_+ - e^{rt_N} C_0^k,$$

where C_0^i denotes the price of the call of maturity t_i and strike K, with

$$S_{t_{k+1}} = S_{t_k} e^{(r - \frac{\sigma^2}{2})(t_{k+1} - t_k) + \sigma\sqrt{(t_{k+1} - t_k)} Z_k}.$$

So that, $X = Z = (Z_1, ..., Z_{250})$ is a Gaussian vector with $d = 250$. Consequently, with our above procedure, θ_n and μ_n are two random vectors of dimension d and we have to control the growth of each component. If one grows too fast and take too high values, it may provides bad performance and bad estimates of both VaR and CVaR. To circumvent this problem, one can reduce the dimension of the problem by choosing the same shift parameters for several dates, *i.e.* for instance

$$\theta_n = (\underbrace{\theta_n^1, .., \theta_n^1}_{10 \text{ times}}, ..., \underbrace{\theta_n^{25}, .., \theta_n^{25}}_{10 \text{ times}}).$$

Now, we can run the IS algorithm for $\theta^1, ..., \theta^{25}$ so that, we have to deal with a procedure in dimension 25. It is sub-optimal with respect to the procedure in dimension 250 but it is more tractable. Another relevant example is a portfolio composed by only one barrier option, for instance a Down & In Call option

$$\varphi(X) = (X_T - K)_+ \mathbf{1}_{\left\{\min_{\{0 \le t \le T\}} X_t \le L\right\}}$$

where the underlying X is a process solution of the path-dependent SDE

$$\mathrm{d}X_t = b(X_t)\ \mathrm{d}t + \sigma(X_t)\ \mathrm{d}W_t,\ X_0 = x \in \mathbb{R}^d, \tag{2.50}$$

$W = (W_t)_{t \in [0,T]}$ being a standard Brownian motion. A naive approach is to discretize (2.50) by an Euler-Maruyama scheme $\bar{X} = (\bar{X}_{t_k})_{k \in \{0,...,n\}}$

$$\bar{X}_{t_{k+1}} = \bar{X}_{t_k} + b(\bar{X}_{t_k})(t_{k+1} - t_k) + \sigma(\bar{X}_{t_k})(W_{t_{k+1}} - W_{t_k}),\ \bar{X}_0 = x_0 \in \mathbb{R}.$$

This kind of approximation is known to be poor for this kind of options. In this case, our IS parameters θ and μ are n-dimensional vectors which correspond to the number of steps in the Euler scheme. Now, if you consider a portfolio composed by several barrier options with different underlyings, the dimension can increase greatly and becomes an important issue, so that our first IS procedure is no longer acceptable and tractable. To overcome this problem, the idea is to shift the entire distribution of X in (2.50) thanks to a Girsanov transformation. This last case is analyzed and investigated in [62]. It can be adapted to our framework (see the next section for further developments)

2.5 Numerical examples

For the sake of simplicity, we focus in this section on the finite dimensional setting and on the computation of the regular CVaR_α ($\Psi \equiv Id$). We first consider the usual Gaussian framework in which the exponential change of measure coincide with the mean translation change of measure. Then we illustrate the algorithm (2.49) in a simple case.

2.5.1 Gaussian framework

In this setting, $X \sim \mathcal{N}(0, I_d)$ and p is given by

$$p(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}}, \quad x \in \mathbb{R}^d,$$

so that (B3) and (B4) are satisfied with $\rho = \frac{1}{2}$ and $b = 2$. In this setting, we already noticed that

$$\begin{aligned} L_3(\xi, \theta, x) &:= \mathbf{1}_{\{\varphi(X-\theta)\geq\xi\}}(2\theta - x), \\ L_4(\xi, \mu, x) &:= \frac{1}{1 + G(-\mu) + \xi^2} \left(\varphi(X-\mu) - \xi\right)_+^2 (2\mu - x). \end{aligned}$$

Moreover, we use a stepwise constant sequence α_n that slowly converges toward α as proposed in Section 3.3. We consider three different portfolios of options (puts and calls) on 1 and 5 underlying assets (except for the last case). In the third case, we study the behaviour of a portfolio composed by a power plant that produces electricity from gas with short positions in calls on electricity. The assets are modeled as geometric Brownian motions for the first two examples. In the third example, the assets (electricity and gas day-ahead prices) are modeled as exponentials of an Ornstein-Uhlenbeck process. This last derivative is priced using an approximation of Margrabe formulae (see e.g. [64]). We assume an annual risk free interest rate of 5%. In each example, we use three different values of the confidence level $\alpha = 95\%,\ 99\%,\ 99.5\%$, which are specified in the Tables. We use the following test portfolios:

Example 1. *Short position in one put with strike $K = 110$ and maturity $T = 1$ year on a stock with initial value $S_0 = 100$ and volatility $\sigma = 20\%$. The loss is given by*

$$\varphi_1(X) := (K - S_T)_+ - e^{rT} P_0$$

with

$$S_T := S_0 e^{\left(\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}X\right)}$$

where $X \sim \mathcal{N}(0,1)$ and P_0 is the initial price at which the put option was sold (it is approximately equal to 10.7). The dimension d of the structural vector X is equal to 1. The numerical results are reported in Table 1.

Example 2. *Short positions in 10 calls and 10 puts on each of the five underlying assets, all options having the same maturity 0.25 year. The strikes are set to 130 for calls, to 110 for puts and the initial spot prices to 120. The underlying assets have a volatility of 20% and are assumed to be uncorrelated. The dimension d of the structural vector X is equal to 5. The numerical results are reported in Table 2.*

Example 3. *Short position in a power plant that produces electricity day by day with a maturity of* $T = 1$ *month and 30 long positions in calls on electricity day-ahead price with the same strike* $K = 60$. *Electricity and gas initial spot prices are* $S_0^e = 40$ \$/MWh *and* $S_0^g = 3$ \$/MMBTU *(*BTU*: British Thermal Unit) with a Heat Rate equals* $h_R = 10$ BTU/kWh *and generation costs* $C = 5$ \$/MWh. *The two spot prices have a correlation of 0.4. The payoff can be written*

$$\varphi_3(X) = \sum_{k=1}^{30} \left(e^{r(T-t_k)} \left(S_{t_k}^e - h_R S_{t_k}^g - C \right)_+ - P_0^c e^{rT} \right) + \left(e^{rT} C_0 - e^{r(T-t_k)} \left(S_{t_k}^e - K \right)_+ \right)$$

where P_0^c *is a proxy of the price of the option on the power plant and is equal to 149.9 and* C_0 *is the price of the call options which is equal to 3.8. This is a sum of spark spread options where we decide to exchange gas and electricity each day during one month. The dimension* d *of the structural vector* X *is equal to 60. The numerical results are reported in Table 3.*

The results displayed in the following tables correspond to the VaR, the CVaR and the variance reduction ratios estimations for both VaR and CVaR procedure using a number of steps specified in the first column, still for the same three levels of α. The variance ratios correspond to the ratio of an estimation of the asymptotic variance using the averaging procedure of (2.16) divided by an estimation of the asymptotic variance using the averaging procedure of (2.36): $\mathrm{VR}_{\mathrm{VaR}}$ corresponds to the variance reduction ratio of the VaR estimate and $\mathrm{VR}_{\mathrm{CVaR}}$ corresponds to the variance reduction ratio of the CVaR estimate. The results emphasize that the IS procedure yields a very significant, sometimes huge variance reduction especially when α is closed to 1.
In the three examples, we define the step sequence by $\gamma_n = \frac{1}{n^\beta + 100}$ where $\beta = \frac{3}{4}$.

Table 2.1: Example 1 Results

Number of steps	α	VaR	CVaR	$\mathrm{VR}_{\mathrm{VaR}}$	$\mathrm{VR}_{\mathrm{CVaR}}$
10 000	95%	24.6	29.9	5.5	30.5
	99%	34.4	37.5	11.1	125.3
	99.5%	37.8	41.4	13.4	192.9
100 000	95%	24.6	30.4	6.6	32.2
	99%	34.2	37.9	11.5	127.9
	99.5%	37.3	40.7	15.1	185
500 000	95%	24.6	30.3	7.7	31.3
	99%	34.2	38	14.6	118.4
	99.5%	37.3	40.5	15.5	184

2.5.2 Esscher transform: the NIG distribution

Now, we consider a simple case of portfolio composed by a long position on a Call option with strike $K = 0.6$ and maturity $T = 1$ year, where the underlying is e^{X_T} ($X_0 = 0$), where X_T is a Normal Inverse Gaussian (NIG) variable, $X_T \sim$

Table 2.2: Example 2 Results

Number of steps	α	VaR	CVaR	VR_{VaR}	VR_{CVaR}
10 000	95%	339	440.5	6.5	14.9
	99%	493.1	561.4	10.1	24.3
	99.5%	540.1	606.4	18.2	37.9
100 000	95%	349.8	439.7	6.7	17
	99%	495.7	563.8	11.3	28.6
	99.5%	544.8	607.8	18.9	40.3
500 000	95%	352.4	439.6	6.8	17.3
	99%	495.2	563	11.1	27.7
	99.5%	545.3	608.4	19.2	37

Table 2.3: Example 3 Results

Number of steps	α	VaR	CVaR	VR_{VaR}	VR_{CVaR}
10 000	95%	115.7	150.5	3.4	6.8
	99%	169.4	196	8.4	12.9
	99.5%	186.3	213.2	13.5	20.3
100 000	95%	118.7	150.5	4.5	8.7
	99%	169.4	195.4	12.6	17.5
	99.5%	188.8	212.9	15.6	29.5
500 000	95%	119.2	150.4	5	9.2
	99%	169.8	195.7	13.1	18.6
	99.5%	188.7	212.8	17	29

$\text{NIG}(\alpha, \beta, \delta, \mu)$, $\alpha > 0$, $|\beta| \leq \alpha$, $\delta > 0$, $\mu \in \mathbb{R}$. Its density is given by

$$p_{X_T}(x,\ \alpha,\ \beta,\ \delta,\ \mu) := \frac{\alpha\delta K_1(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{\pi\sqrt{\delta^2 + (x-\mu)^2}} e^{\delta\gamma + \beta(x-\mu)},$$

where K_1 is a modified Bessel function of the second kind and $\gamma = \sqrt{\alpha^2 - \beta^2}$. Note that the generating function of the NIG distribution is given by

$$\psi(\theta) = \mu\theta + \delta(\gamma - \sqrt{\alpha^2 - (\beta + \theta)^2}),$$

and is not well defined for every $\theta \in \mathbb{R}$, so that we change the algorithm parametrization (see section 4.3 of [62]). The loss of the portfolio can be written $L = \varphi_4(X_T) = 50(e^{X_T} - K)_+ - e^{rT}C_0$. Note that the price C_0 is computed by a crude Monte Carlo and is approximately equal to 42. The parameters of the NIG random variable X_T are $\alpha = 2.0,\ \beta = 0.2,\ \delta = 0.8,\ \mu = 0.04$. We want to compare the variance reduction achieved by the translation of the mean (see section 3.1) and the one achieved by the Esscher Transform (see section 4.1). In the Robbins-Monro procedure, we define the step sequence by $\gamma_n = \frac{1}{n^\beta + 100}$ where $\beta = \frac{3}{4}$.

▷ *Translation case.* The functions L_3 and L_4 of the IS procedure are defined by:

$$\begin{aligned} L_3(\xi,\theta,X) &:= e^{-2|\theta|}\mathbf{1}_{\varphi(X-\theta)}\frac{p'(X-2\theta)}{p(X)}\left(\frac{p(X-\theta)}{p(X-2\theta)}\right)^2, \\ L_4(\xi,\mu,X) &:= \frac{e^{-2|\mu|}}{1+G(-\mu)+\xi^2}(\varphi(X-\mu)-\xi)_+^2\frac{p'(X-2\mu)}{p(X)}\left(\frac{p(X-\mu)}{p(X-2\mu)}\right)^2, \end{aligned}$$

where p' is easily obtained using the relation on the modified Bessel function $K_1'(x) = \frac{1}{x}K_1(x) - K_2(x)$.

▷ *Esscher Transform.* In this approach, the functions L_3 and L_4 are defined by

$$\begin{aligned} L_3(\xi,\theta,X) &:= \mathbf{1}_{\varphi(X^{(-\theta)})\geq\xi}(\nabla\psi(\theta) - X^{(-\theta)}), \\ L_4(\xi,\mu,X) &:= \frac{e^{-|\mu|}}{1+\xi^2}(\varphi(X^{(-\mu)})-\xi)_+^2(\nabla\psi(\mu) - X^{(-\mu)}), \end{aligned}$$

where $X^{(\pm\theta)} \sim \mathrm{NIG}(\alpha, \beta \pm \theta, \delta, \mu)$.
Table 4 compares the variance reduction ratios of the VaR_α and CVaR_α algorithms achieved by the translation of the mean (VR^{tr}_{VaR} and VR^{tr}_{CVaR}) and the one achieved by the Esscher Transform (VR^{es}_{VaR} and VR^{es}_{CVaR}).

Table 2.4: Example 4 Results

Number of steps	α	VaR	CVaR	$\mathrm{VR}^{tr}_{\mathrm{VaR}}$	$\mathrm{VR}^{tr}_{\mathrm{CVaR}}$	$\mathrm{VR}^{es}_{\mathrm{VaR}}$	$\mathrm{VR}^{es}_{\mathrm{CVaR}}$
10 000	95%	85.8	215.7	5	10	4.2	58.8
	99%	217	518	6	12	8	60
	99.5%	304	748	8	25	8.9	110
100 000	95%	87.2	215.1	5	12	4.5	60
	99%	218	521	5	12	8.2	70
	99.5%	303.5	747.8	7	30	12	100
500 000	95%	87.9	215.6	5	9	5	57
	99%	227	518.9	5.5	11.8	11.5	68
	99.5%	312.8	741.8	6	31	10	123

The IS procedure is very efficient when $\mathbb{P}(\varphi(X) \geq \xi^*) = 1-\alpha$ is close to zero and becomes more and more efficient as α grows to 1. Even for the complex portfolio considered in Example 3, where X is a Gaussian vector with $d = 60$, it is possible to achieve a great variance reduction for both VaR_α and CVaR_α.

We observed that IS based on Esscher transform is well adapted to distributions with heavy tails (*i.e.* heavier tails than the normal distribution). It is therefore suitable when large values are more frequent than for the normal distribution, as it is the case when the vector X is a NIG random variable. Indeed, in this setting, the IS parameters modify the parameter β which controls the asymmetric shape of the NIG distribution. We think that the IS procedure by Esscher transform outperforms the IS procedure by mean translation when the IS parameter impacts on the symmetry of the distribution.

Concluding Remarks

In this article, we propose a recursive procedure to compute efficiently the Value-at-Risk and the Conditional Value-at-Risk using the same innovation for both procedures. In our approach, for a given risk level α, the VaR_α and the CVaR_α are estimated simultaneously by a regular RM algorithm. Ruppert and Polyak's averaging principle provides an asymptotically efficient procedure. The estimates satisfy a Gaussian CLT. However, due to the slow convergence of the global procedure since we are interested in rare events, the regular version of this algorithm cannot be used in practice. To speed-up and thus greatly reduce the number of scenarios, we devise an unconstrained adaptive IS procedure. The resulting procedure provides estimates that satisfy a CLT with minimal variances. To optimize the move to the critical risk area, the risk level α can be temporarily replaced by a slowly increasing level α_n (stepwise constant in practice) converging to α. This produces a VaR *companion procedure* $(\hat{\xi}_n)_{n\geq 1}$ that controls the IS change of measure parameters $(\hat{\theta}_n, \hat{\mu}_n)$. Numerically speaking, the resulting procedure converges efficiently and can drastically reduce the variance. It is possible to extend the methods to portfolio whose losses depend on a general diffusion process, using Girsanov transform to introduce a potentially infinite dimensional variance reducer. Finally, we aim at extending the method by implementing low-discrepancy sequences in our procedure instead of pseudo-random numbers. Preliminary numerical experiments showed a significant improvement of the convergence rate. This also raises interesting theoretical problems. See [58] for some first theoretical results in that direction in a one-dimensional framework and chapter 4 for further developments in higher dimensional setting.

Chapter 3

VaR-CVaR algorithm with unconstrained importance sampling: the infinite dimensional setting

This work was submitted to the Journal *Monte Carlo Methods and Applications*.

Abstract: In this paper, we extend the method of finding Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) by stochastic approximation and adaptive importance sampling, as introduced in [6], to a large class of portfolio losses that depend on a path-dependent diffusion like processes. In this context, we show how to devise a regular Robbins-Monro algorithm to approximate the optimal importance sampling (I.S.) process, which somehow follows the new approach introduced in [62]. Like in the finite dimensional framework, the convergence rate of the resulting algorithm to its target satisfies a Gaussian Central Limit Theorem with optimal rate and minimal variance.

Numerous simulations are provided to assert its efficiency. In particular, we carry out a comparison with the finite dimensional procedure on an energy related portfolio.

Keywords: VaR, CVaR, Stochastic Approximation, Robbins-Monro algorithm, Importance Sampling, Girsanov Transform.

3.1 Introduction

Measuring the risk of a financial portfolio has become a key problem for financial institutions and energy companies. The Value-at-Risk (VaR) is widely used in this context to summarize extreme losses potentially faced by an agent that holds the considered portfolio. However, it has recognized several limitations. It lacks subadditivity and convexity so that it can discourage diversification. Moreover, it does not take into account the magnitude and the distribution of the loss beyond the VaR. Following [3], it does not satisfy the set of requirements of coherent risk measure. An alternative risk measure to VaR that has many attractive properties, including subadditivity and convexity, is the Conditional Value-at-Risk (CVaR).

Several methods have been proposed to compute VaR using historical or Monte Carlo simulation. The most commonly used method is the inversion of the empirical loss distribution function. Approximations of the portfolio loss distribution known as "Delta-Gamma approximation" relying on a linear or quadratic expansion are also used, see e.g. [17], [39], [38]. Such approximations are non longer tractable and acceptable when considering portfolios with long maturity (1 year up to 10 years) or when the loss is a functional of a general path-dependent Stochastic Differential Equation (S.D.E.).

In the context of portfolio optimization to reduce the CVaR, it is demonstrated in [77] that it is possible to approximate both VaR and CVaR (optimize CVaR) by solving (an approximation of) a convex optimization problem which reduces to a linear programming problem. However, the linear programming problem is subject to huge number of linear constraints so that it remains difficult to handle in practice.

Recently, in [6], we proposed a method for computing both VaR and CVaR based on a stochastic approximation algorithm derived from the fact that they are the solution and the value of the same convex optimization problem as demonstrated in [77]. In [6], it is shown that the convergence rate of this algorithm to its target satisfies a Gaussian Central Limit Theorem under standard assumptions. However, the original procedure converges slowly as soon as the confidence level is closed to 1, as it is often the case in practice. Actually, in this kind of problem, we are interested by events that are observed with a very low probability (usually less that 5%). Thus, as a necessary improvement, an adaptive variance reduction procedure based on a recursive importance sampling (I.S.) stochastic approximation algorithm was proposed. Originally introduced by [1] in the gaussian case, it has been revisited recently in [62] in a more general framework. The I.S. procedure combined with the VaR-CVaR procedure in [6] follows this new approach. However, as noticed in [6], when we have to sample from a high dimensional structural random vector X ($X \in \mathbb{R}^d$ with $d \geq 50$) to simulate the loss L, as it is often the case in financial institutions or energy companies, we have to optimize two vectors of dimension d and thus control the growth of each component. Actually, in most cases, X is a vector of Brownian increments related to the Euler scheme of a diffusion so that the dimension of X is often greater than 100 especially if the portfolio is composed of several market prices. This is no longer tractable.

In this paper, we show that it is possible to extend the algorithm investigated in [6] to the case where the loss is a functional of a path-dependent diffusion X. To be more precise, we want now to compute both VaR and CVaR of $L := \varphi(X)$ where φ is a functional defined on the space $\mathcal{C}([0,T],\mathbb{R}^d)$ of continuous functions

defined on $[0,T]$ and to devise a Robbins-Monro I.S. algorithm based on the approach investigated in [62]. Thus, φ is the function describing the composition of the portfolio which remains fixed and X models the market prices over the time interval; thus we do not need to specify the dynamics of the market prices and only rely on the fact that it is possible to sample from the distribution of X. Thus, we consider a d-dimensional Itô process X solution to the stochastic differential equation (S.D.E.)

$$\mathrm{d}X_t = b(t, X^t)\,\mathrm{d}t + \sigma(t, X^t)\,\mathrm{d}W_t,\ X_0 = x \in \mathbb{R}^d, \qquad (E_{b,\sigma})$$

$W = (W_t)_{t\in[0,T]}$ being a q-dimensional standard Brownian motion and where $X^t := (X_{t\wedge s})_{s\in[0,T]}$ is the stopped process at time t, $b : [0,T]\times\mathcal{C}([0,T],\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : [0,T]\times \mathcal{C}([0,T],\mathbb{R}^d) \to \mathcal{M}(d,q)$ are measurable with respect to the canonical predictable $\sigma-$field on $[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)$. The VaR at level $\alpha \in (0,1)$ is the lowest α-quantile of the distribution $\varphi(X)$ *i.e.*:

$$\mathrm{VaR}_\alpha(\varphi(X)) := \inf\left\{\xi \mid \mathbb{P}\left(\varphi(X) \le \xi\right) \ge \alpha\right\}.$$

We assume that the distribution function of $\varphi(X)$ is continuous (*i.e.* without atoms) so that it is the lowest solution of the equation:

$$\mathbb{P}(\varphi(X) \le \xi) = \alpha.$$

If the distribution function is strictly increasing, the solution of the above equation is unique, otherwise, there may be more than one solution. In fact, in what follows, we will consider that *any* solution of the previous equation is the VaR. Like in the static framework, we also introduce the Ψ-*Conditional Value-at-Risk* (Ψ-CVaR) (at level α). As soon as $\Psi(\varphi(X)) \in L^1(\mathbb{P})$, it is defined by:

$$\Psi\text{-CVaR}_\alpha(\varphi(X)) := \mathbb{E}\left[\Psi(\varphi(X))|\varphi(X) \ge \mathrm{VaR}_\alpha(\varphi(X))\right]. \qquad (3.1)$$

When $\Psi \equiv Id$ and $\varphi(X) \in L^1(\mathbb{P})$, (3.1) is the regular CVaR of $\varphi(X)$.

This paper is organized as follows. Section 2 is devoted to some background on the finite dimensional setting where we briefly recall the main tools and assumptions used to devise the VaR-CVaR investigated in [6]. In Section 3, we introduce the functional version of the algorithm where the I.S. procedure is based on the Girsanov theorem. In Section 4, numerical experiments are carried out on several portfolios to emphasize its efficiency.

Notations: • $|.|$ will denote the canonical Euclidean norm on $\mathbb{R}^d$ and $\langle .,.\rangle$ will denote the canonical inner product.

• $\xrightarrow{\mathcal{L}}$ will denote the convergence in distribution and $\xrightarrow{a.s.}$ will denote the almost sure convergence.

• $\|f\|_{L^2_{T,p}} := \left(\int_0^T \left(f_1^2(t) + \cdots + f_p^2(t)\right)\mathrm{d}t\right)^{\frac{1}{2}}$ if $f = (f_1,\ldots,f_p)$ is an $\mathbb{R}^p$-valued (class of) Borel function(s).

• $x_+ := \max(0,x)$ will denote the positive part function.

3.2 Some background: the finite-dimensional setting

In [6], we proposed a Robbins-Monro (R.M.) algorithm to compute the VaR_α and the Ψ-CVaR_α, at a given confidence level $\alpha \in (0,1)$, for a portfolio whose loss L can be written $L = \varphi(X)$, where $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, |.|)$ and $\varphi,\ \Psi : \mathbb{R}^d \to \mathbb{R}$ are two Borel functions ($\Psi(\varphi(X))$ is integrable). We proved that the VaR_α and the Ψ-CVaR_α can be estimated as the limit of the following algorithm

$$Z_n = Z_{n-1} - \gamma_n H(Z_{n-1}, X_n),\ n \geq 1 \tag{3.2}$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence of random vectors having the same distribution of X, $Z_0 = (\xi_0, C_0)$ is a random vector independent of $(X_n)_{n\geq 1}$ such that $\mathbb{E}[|\xi_0|^2] < +\infty$ with $H(z,x) := (H_1(\xi,x), H_2(\xi,C,x))$ and for all $x \in \mathbb{R}^d$

$$H_1(\xi,x) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\varphi(x)\geq\xi}$$
$$H_2(\xi,C,x) = C - \xi - \frac{1}{1-\alpha}(\Psi(\varphi(x)) - \xi)\mathbf{1}_{\varphi(x)\geq\xi}.$$

To obtain the *a.s.* convergence and the weak convergence rate of the procedure, we introduced the quite standard step assumption

$$\sum_{n\geq 1}\gamma_n = +\infty \text{ and } \sum_{n\geq 1}\gamma_n^2 < +\infty. \tag{A1}$$

and the following additional assumption on the distributions of $\varphi(X)$ and $\Psi(\varphi(X))$

$$\varphi(X) \text{ is continuous and increasing, and } \Psi(\varphi(X)) \in L^{2a}(\mathbb{P}) \text{ for } a > 0. \tag{$A2)_a$}$$

The two main results of [6] are the two following propositions.

Proposition 3.2.1. *(a.s. convergence) Suppose that assumptions* $(A1)$ *and* $(A2)_1$ *are satisfied. Then,* (3.2) *a.s. converges toward* $z^* := (\xi_\alpha^*, C_\alpha^*)$ *where* ξ_α^* *is a random variable taking its value in* $\{\xi \in \mathbb{R} \mid \mathbb{P}(\varphi(X) \leq \xi) = \alpha\}$ *and* $C_\alpha^* = \Psi$-CVaR_α.

Proposition 3.2.2. *(Convergence rate) Suppose that assumptions* $(A1)$ *and* $(A2)_a$ *are satisfied for some* $a > 1$. *Assume that the distribution has a positive probability density* $f_{\varphi(X)}$ *on its support (which implies that* $VaR_\alpha(\varphi(X))$ *is unique so that the sequence* $(Z_n)_{n\geq 1}$ *a.s. converges to its single target* $z_\alpha^* := (\mathrm{VaR}_\alpha(\varphi(X)), \Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X)))$. *Let* $(\bar{Z}_n)_{n\geq 1}$ *be the Cesàro mean of the sequence* $(Z_n)_{n\geq 1}$, i.e. *the sequence defined by*

$$\bar{Z}_n := \frac{Z_0 + \cdots + Z_{n-1}}{n},\ n \geq 1,$$

which converges to z_α^*. *If the step sequence is defined by* $\gamma_n := \frac{c}{n^\beta}$ *with* $\frac{1}{2} < \beta < 1$, *the Ruppert and Polyak's Averaging Principle ensures that*

$$\sqrt{n}\left(\bar{Z}_n - z^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma) \qquad \text{as } n \to +\infty$$

where Σ *is given by*

$$\begin{pmatrix} \frac{\alpha(1-\alpha)}{f_{\varphi(X)}^2(\xi^*)} & \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[(\Psi(\varphi(X)) - \xi^*)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\ \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[(\Psi(\varphi(X)) - \xi^*)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X)) - \xi^*)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right) \end{pmatrix}.$$

For more details about this algorithm, we refer to [6].

This procedure is known to converge slowly owing to the fact that $\mathbb{P}(\varphi(X) > \xi_\alpha^*) = 1-\alpha$ is close to zero as the confidence level α is often close to 1 in practical applications (otherwise the procedure behaves well), meaning that we observe fewer and fewer simulations for which $\varphi(X_k) > \xi_{k-1}$ as the algorithm evolves. Moreover, practitioners usually deal with huge portfolio so that the evaluation step of $\varphi(X)$ may consume a lot of time. Consequently, as a crucial improvement, we combined the VaR-CVaR algorithm with a recursive variance reduction method based on an adaptive I.S. algorithm. The aim of this procedure is to modify the distribution of X in order to minimize the asymptotic variance of the two components in the above CLT, the asymptotic variance of the VaR_α algorithm

$$\frac{\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X)\geq\xi_\alpha^*\}}\right)}{f_{\varphi(X)}^2} = \frac{\alpha(1-\alpha)}{f_{\varphi(X)}^2}$$

and the asymptotic variance of Ψ-CVaR_α algorithm, namely

$$\frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X)) - \xi_\alpha^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi_\alpha^*\}}\right).$$

We investigated the translation of mean and the exponential change of measure (also called the Esscher transform). In both cases, it consists in approximating the optimal parameters θ_α^* (for the VaR algorithm), μ_α^* (for the Ψ-CVaR algorithm) by two R.M. procedures in $\mathbb{R}^d$, namely:

$$\theta_n = \theta_{n-1} - \gamma_n H_3(\xi_{n-1}, \theta_{n-1}, X_n),\ \theta_0 \in \mathbb{R}^d$$
$$\mu_n = \mu_{n-1} - \gamma_n H_4(\xi_{n-1}, \mu_{n-1}, X_n),\ \mu_0 \in \mathbb{R}^d$$

where ξ_n is the approximation at step n of the VaR algorithm modified by the I.S. procedure (for more details, we refer to [6]). Generally, in the energy sector as in financial institutions, we want to estimate the VaR and the CVaR on a huge portfolio, so that the dimension of the random vector X can be very large (hundreds up to thousands for portfolio with a long maturity). As already noticed in the introduction, we have to control the growth of each component of $(\theta_n, \mu_n)_{n\geq 1}$ to ensure its *a.s.* convergence. If one component remains "stuck" at the beginning of the algorithm, this may provide bad performance and bad estimates of both VaR and CVaR. This version of the I.S. algorithm is no longer tractable in a high dimensional setting.

3.3 VaR-CVaR algorithm for diffusions

3.3.1 Framework and practical results

Applied to the computation of the mean $\mathbb{E}[F(X)]$ ($F \in L^2(\mathbb{P}_X)$ with $\mathbb{P}(F(X) \neq 0) > 0$), it is shown in [62] that it is possible to circumvent the problem of the dimension by extending the adaptive I.S. algorithm to infinite dimensional setting (where X is a path-dependent diffusion like processes). In this framework, the loss of the portfolio is related to a diffusion X solution of $(E_{b,\sigma,W})$, the loss functional φ is defined on the

space $\mathcal{C}([0,T],\mathbb{R}^d)$ of continuous functions defined on $[0,T]$ and the I.S. change of measure will be based on the Girsanov transform. Under the following assumption

$$\left\{\begin{array}{l}(i)\ \ b(.,0) \text{ and } \sigma(.,0) \text{ are continuous,}\\ \\ (ii)\ \ \forall t\in[0,T], \forall x,y\in\mathcal{C}([0,T],\mathbb{R}^d),\ |b(t,y)-b(t,x)|\\ \qquad\qquad\qquad\qquad +||\sigma(t,y)-\sigma(t,x)||\leq C_{b,\sigma}||x-y||_\infty.\end{array}\right. \tag{$\mathcal{H}_{b,\sigma}$}$$

strong existence and uniqueness of solutions for $(E_{b,\sigma,W})$ can be proved (for more details, see [79]).
If $b(t,x^t)=\beta(t,x(\underline{t}))$ and $\sigma(t,x^t)=\vartheta(t,x(\underline{t}))$ for every $x\in\mathcal{C}([0,T],\mathbb{R}^d)$, where $\underline{t}:=\left\lfloor\frac{tn}{T}\right\rfloor\frac{T}{N}$, then X is *the continuous Euler scheme with step* $\frac{T}{N}$ of $(E_{b,\sigma,W})$ with drift β and diffusion coefficient ϑ.

Let f be a fixed borel bounded functional on $\mathcal{C}([0,T],\mathbb{R}^d)$ with values in $\mathcal{M}(q,p)$. Then a Girsanov transform yields that for every $\theta\in L^2_{T,p}:=L^2([0,T],\mathbb{R}^p)$,

$$\mathbb{E}\left[F\left(X\right)\right]=\mathbb{E}\left[F\left(X^{(\theta)}\right)e^{-\int_0^T\left\langle f(X^{(\theta),s})\theta_s,dW_s\right\rangle-\frac{1}{2}||f(X^{(\theta),\cdot})\theta_\cdot||^2_{L^2_{T,q}}}\right]$$

where $X^{(\theta)}$ denotes the strong solution to $(E_{b+\sigma f\theta,\sigma})$. As we already noticed in [6], applying this idea to the computation of both VaR and CVaR amounts to translate the distribution of X in order to minimize the asymptotic variance of the VaR_α $\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X)>\xi^*\}}\right)$ and of the CVaR_α algorithm $\mathrm{Var}\left((\Psi(\varphi(X))-\xi^*)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right)$. Let f_1 and f_2 be two Borel functions, $f_1,f_2:\mathcal{C}([0,T],\mathbb{R}^d)\to\mathcal{M}(q,p)$. The two functionals to be minimized are

$$\min_{\theta\in L^2_{T,p}}Q_1(\theta,\xi^*_\alpha)\ \text{ and }\ \min_{\mu\in L^2_{T,p}}Q_2(\mu,\xi^*_\alpha)$$

where

$$Q_1(\theta,\xi^*_\alpha)=\mathbb{E}\left[\mathbf{1}_{\left\{\varphi(X^{(\theta)})>\xi^*_\alpha\right\}}e^{-2\int_0^T\left\langle f_1(X^{(\theta),s})\theta_s,dW_s\right\rangle-||f_1(X^{(\theta),\cdot})\theta_\cdot||^2_{L^2_{T,q}}}\right]$$

and,

$$Q_2(\mu,\xi^*_\alpha)=\mathbb{E}\left[(\Psi(\varphi(X^{(\mu)}))-\xi^*_\alpha)^2\mathbf{1}_{\left\{\varphi(X^{(\mu)})>\xi^*_\alpha\right\}}e^{-2\int_0^T\left\langle f_2(X^{(\mu),s})\mu_s,dW_s\right\rangle-\left\|f_2\left(X^{(\mu),\cdot}\right)\mu_\cdot\right\|^2_{L^2_{T,q}}}\right].$$

In practice, we will only minimize Q_1 and Q_2 over a *finite dimensional subspace* $E=\mathrm{span}\{e_1,...,e_m\}\subset L^2_{T,p}$. Using another Girsanov Transform (see Lemma 3 in [62]) yields

$$Q_1(\theta,\xi^*_\alpha)=\mathbb{E}\left[\mathbf{1}_{\{\varphi(X)>\xi^*_\alpha\}}e^{-\int_0^T\left\langle f_1(X^{(\theta)})\theta_s,dW_s\right\rangle+\frac{1}{2}||f_1(X^{(\theta),\cdot})\theta_\cdot||^2_{L^2_{T,q}}}\right],$$

and

$$Q_2(\mu,\xi^*_\alpha)=\mathbb{E}\left[(\Psi(\varphi(X))-\xi^*_\alpha)^2\mathbf{1}_{\{\varphi(X)>\xi^*_\alpha\}}e^{-\int_0^T\left\langle f_2(X^{(\mu),s})\mu_s,dW_s\right\rangle+\frac{1}{2}||f_2(X^{(\mu),\cdot})\mu_\cdot||^2_{L^2_{T,q}}}\right].$$

Now we are in position to ensure under some assumptions that Q_1 (resp. Q_2) is finite, log-convex, and goes to infinity as $\|\theta\|_{L^2_{T,p}}$ (resp. $\|\mu\|_{L^2_{T,p}}$) goes to infinity so that Argmin Q_1 and Argmin Q_2 are *non empty*.

Moreover, we will show that, $Q_1(.,\xi^*_\alpha)$ and $Q_2(.,\xi^*_\alpha)$ are differentiable thus Argmin $Q_1 = \{\nabla_\theta Q_1(.,\xi^*_\alpha) = 0\}$ and Argmin $Q_2 = \{\nabla_\mu Q_2(.,\xi^*_\alpha) = 0\}$. This will allow us to derive a stochastic gradient algorithm. For that, we need to introduce the following assumption where Argmin $V = \{\xi \mid \mathbb{P}(\varphi(X) > \xi) = 1-\alpha\}$

$$\mathbb{P}\left((\Psi(\varphi(X)) - \xi^*_\alpha)^2 \mathbf{1}_{\{\varphi(X)>\xi^*_\alpha\}} > 0\right) \; > \; 0, \; \forall \xi^*_\alpha \in \text{Argmin } V. \tag{A3}$$

Proposition 3.3.1. *Suppose that $(A2)_a$ holds for some $a > 1$ as well as (A3) and $(\mathcal{H}_{b,\sigma})$. Then, for every $\xi \in \mathbb{R}$, $Q_1(.,\xi)$ and $Q_2(.,\xi)$ are finite on $L^2_{T,p}$ and* log*-convex.*

1. *Assume that the bounded matrix-valued Borel functions f_1 and f_2 satisfy that $f_1(X^s)$ (resp. $f_2(X^s)$) has a non-atomic kernel on the event $\{\varphi(X) > \xi^*_\alpha\}$ (resp. on the event $\left\{(\Psi(\varphi(X)) - \xi^*_\alpha)^2 \mathbf{1}_{\{\varphi(X)>\xi^*_\alpha\}} > 0\right\}$)* i.e.

$$\begin{aligned}
&\mathbb{P}\left(\left\{\exists\, \theta \in L^2_{T,p}\backslash\{0\} \;\; s.t.\; \theta(s) \in \; Ker f_1(X^s)\; ds - a.e \text{ and } \varphi(X) > \xi^*_\alpha\right\}\right) = 0 \\
&\mathbb{P}\left(\left\{\exists\, \theta \in L^2_{T,p}\backslash\{0\} \;\; s.t.\; \theta(s) \in \; Ker f_2(X^s)\; ds - a.e \right.\right. \\
&\qquad\qquad \left.\left. \text{and } \; (\Psi(\varphi(X)) - \xi^*_\alpha)^2 \mathbf{1}_{\{\varphi(X)>\xi^*_\alpha\}} > 0\right\}\right) = 0
\end{aligned} \tag{3.3}$$

   then for every finite dimensional subspace $E \subset L^2_{T,p}$,

$$\lim_{||\theta||_{L^2_{T,p}} \to +\infty,\; \theta\in E} Q_1(\theta,\xi^*_\alpha) = +\infty \quad \text{and} \quad \lim_{||\mu||_{L^2_{T,p}} \to +\infty,\; \mu\in E} Q_2(\mu,\xi^*_\alpha) = +\infty.$$

2. *For every $\xi \in \mathbb{R}$, the functions $Q_1(.,\xi)$ and $Q_2(.,\xi)$ are differentiable at every $\theta,\ \mu \in L^2_{T,p}$ and the differentials $DQ_1(\theta,\xi)$ and $DQ_2(\mu,\xi)$ are characterized on every $\zeta \in L^2_{T,p}$ by*

$$\begin{aligned}
\langle DQ_1(\theta,\xi),\ \zeta\rangle_{L^2_{T,p}} = \mathbb{E}\Bigg[&\mathbf{1}_{\left\{\varphi(X^{(-\theta)})>\xi\right\}} e^{||f_1(X^{(-\theta),\cdot})\theta_\cdot||_{L^2_{T,p}}} \\
&\times \left(2\left\langle f_1\left(X^{(-\theta),\cdot}\right)\theta_\cdot, f_1(X^{(-\theta),\cdot})\zeta_\cdot\right\rangle_{L^2_{T,p}} - \int_0^T \left\langle f_1(X^{(-\theta),s})\zeta_s, dW_s\right\rangle\right)\Bigg]
\end{aligned}$$

   and,

$$\begin{aligned}
\langle DQ_2(\mu,\xi),\ \zeta\rangle_{L^2_{T,p}} = \mathbb{E}\Bigg[&\left(\Psi(\varphi(X^{(-\mu)})) - \xi\right)^2 \mathbf{1}_{\left\{\varphi(X^{(-\mu)})>\xi\right\}} e^{||f_2(X^{(-\mu),\cdot})\mu_\cdot||_{L^2_{T,p}}} \\
&\times \left(2\left\langle f_2\left(X^{(-\mu),\cdot}\right)\mu_\cdot, f_2(X^{(-\mu),\cdot})\zeta_\cdot\right\rangle_{L^2_{T,p}} - \int_0^T \left\langle f_2(X^{(-\mu),s})\zeta_s, dW_s\right\rangle\right)\Bigg].
\end{aligned}$$

Proof. The proof of those results follows the line of the proof of Proposition 4 in [62]. □

For practical implementations, the first conclusion is the only result of interest. It ensures that

$$\mathcal{T}^* := \left\{(\xi,\theta,\mu) \mid \xi \in \text{Argmin} V,\ \theta \in \text{Argmin}_{|E} Q_1(.,\xi),\ \mu \in \text{Argmin}_{|E} Q_2(.,\xi)\right\} \neq \emptyset.$$

3.3.2 The resulting algorithm

Following the design of the algorithm in [62], for practical implementation, we consider finite dimensional spaces E of $L^2_{T,p}$. Thanks to the result above, we know that the restriction of $Q_1(.,\xi^*_\alpha)$ (resp. $Q_2(.,\xi^*_\alpha)$) on E attains a minimum $\theta^*_{\alpha,E}$ (resp. $\mu^*_{\alpha,E}$) which will be the target of our adaptive I.S. algorithm. Like for the static framework, the *a.s.* convergence of the global algorithm (composed of the VaR algorithm and the two I.S. procedures) toward $z^*_\alpha := (\xi^*_\alpha, \theta^*_\alpha, \mu^*_\alpha) \in \mathcal{T}^*$ is ensured by the Robbins-Monro Theorem (in fact, we will rely on a slight extension which takes into account the case of non-uniqueness of the target, see e.g. [6]) applied to the Lyapunov function $L(z) := \|z - z^*_\alpha\|^2$. Then, the *a.s.* convergence of the Ψ-CVaR$_\alpha$ algorithm follows. The I.S. algorithm is based on the above representation of the two differential. However, first we need to control the discrepancy between X and $X^{(-\theta)}$ (resp. $X^{(-\mu)}$) (this is a general result about SDE, see e.g. Lemma 4 in [62])

Lemma 3.3.2. *Assume that $(\mathcal{H}_{b,\sigma})$ holds. Let f be a bounded Borel $\mathcal{M}(q,p)$-valued function defined on $\mathcal{C}([0,T],\mathbb{R}^d)$, let $\theta \in L^2_{T,p}$ and let X and $X^{(\theta)}$ denote strong solutions of $E_{b,\sigma,W}$ and $E_{b+\sigma f\theta,\sigma,W}$ driven by the same Brownian motion W. Then, for every $r \geq 1$, there exists a real constant $C_{b,\sigma} > 0$ such that*

$$\left\| \sup_{t\in[0,T]} \left| X_t - X^{(\theta)}_t \right| \right\|_r \leq C_{b,\sigma} e^{C_{b,\sigma}T} \left\| \int_0^T \left| \sigma\left(s, X^{(\theta),s}\right) f\left(X^{(\theta),s}\right) \theta_s \right| ds \right\|_r .$$

Remark 1. *Note that it is crucial that we have for all Brownian motion W a strong solution $X^{(-\theta)}$. This is case if X is the Euler scheme of a diffusion or if the driver f is Lipschitz in space.*

Like for the static framework, we need to control the growth of the function $\Psi o\varphi$. So that, we are led to introduce the following assumption

$$\exists \lambda > 0,\ \forall x \in \mathcal{C}([0,T],\mathbb{R}^d), \quad |\Psi(\varphi(x))| \leq C_{\Psi,\varphi}\left(1 + \|x\|^\lambda_\infty\right) \qquad (G_{\Psi,\varphi,\lambda})$$

which induces that $\Psi(\varphi(X)) \in L^r(\mathbb{P})$ for every $r > 0$. Now given a finite dimensional subspace E of $L^2_{T,p}$ spanned by an orthonormal basis $(e_1, ..., e_m)$ and two bounded Borels $\mathcal{M}_{q,p}$-valued functions $f_1,\ f_2$ (with $p \geq 1$), we are lead to define two procedures by

$$\begin{cases} \theta_n = \theta_{n-1} - \gamma_n H_3\left(\theta_{n-1}, \xi_{n-1}, X^{(-\theta_{n-1})}, W^{(n)}\right),\ \theta_0 \in E, \\ \mu_n = \mu_{n-1} - \gamma_n H_{4,\lambda,\eta}\left(\mu_{n-1}, \xi_{n-1}, X^{(-\mu_{n-1})}, W^{(n)}\right),\ \mu_0 \in E, \end{cases} \qquad (3.4)$$

where $(\gamma_n)_{n\geq 1}$ satisfies (A1), $(W^{(n)})_{n\geq 1}$ is a sequence of independent Brownian motions, $X^{(-\theta_{n-1})} = F(-\theta_{n-1}, W^{(n)})$ $\left(\text{resp. } X^{(-\mu_{n-1})} = F(-\mu_{n-1}, W^{(n)})\right)$ is a strong solution to $\left(E_{b-\sigma f_1\theta_{n-1},W^{(n)}}\right)$ $\left(\text{ resp. } \left(E_{b-\sigma f_2\mu_{n-1},W^{(n)}}\right)\right)$, $(\xi_n)_{n\geq 0}$ is the VaR algorithm (modified by the I.S. procedure) defined by

$$\xi_n = \xi_{n-1} - \gamma_n H_1\left(\xi_{n-1}, \theta_{n-1}, X^{(\theta_{n-1})}, W^{(n)}\right),\ \xi_0 \in L^2(\mathbb{P}), \qquad (3.5)$$

and for every standard Brownian motion W and every $\mathcal{F}^W_t$-adapted $\mathbb{R}^p$-valued process $\zeta = (\zeta_t)_{t\in[0,T]}$, H_1, H_3 and $H_{4,\lambda,\eta}$ are defined by

$$H_1(\xi,\theta,\zeta,W) = e^{-\frac{1}{2}||f_1||_\infty||\theta||^2_{L^2_{T,q}}}\left(1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(\zeta)>\xi\}} e^{-\int_0^T \langle f_1(\zeta^s)\theta_s, dW_s\rangle - \frac{1}{2}||f_1(\zeta^s)\theta_\cdot.||^2_{L^2_{T,q}}}\right),$$

and

$$\begin{cases} \langle H_3(\theta,\xi,\zeta,W),e_i\rangle_{L^2_{T,p}} = \mathbf{1}_{\{\varphi(\zeta)>\xi\}} e^{\|f_1(\zeta^\cdot)\theta_\cdot\|^2_{L^2_{T,q}}} \Big(2\langle f_1(\zeta^\cdot)\theta_\cdot, f_1(\zeta^\cdot)e_i\rangle_{L^2_{T,q}} \\ \qquad\qquad - \int_0^T \langle f_1(\zeta^s)e_i(s), dW_s\rangle\Big), \\ \langle H_{4,\lambda,\eta}(\mu,\xi,\zeta,W),e_i\rangle_{L^2_{T,p}} = C_{\lambda,\eta}(\mu,\zeta)\left(\Psi(\varphi(\zeta))-\xi\right)^2 \mathbf{1}_{\{\varphi(\zeta)>\xi\}} e^{\|f_2(\zeta^\cdot)\mu_\cdot\|^2_{L^2_{T,q}}} \\ \qquad\qquad \times\Big(2\langle f_2(\zeta^\cdot)\mu_\cdot, f_2(\zeta^\cdot)e_i\rangle_{L^2_{T,q}} - \int_0^T \langle f_2(\zeta^s)e_i(s), dW_s\rangle\Big), \end{cases}$$

$i=1,...,m$ where for $\eta>0$,

$$C_{\lambda,\eta}(\mu) = \begin{cases} \frac{e^{-\|f_2\|_\infty \|\mu\|^2_{L^2_{T,p}}}}{1+\xi^2+\|\mu\|^{2\lambda+\eta}_{L^2_{T,p}}} & if\ \sigma\ is\ bounded, \\ \frac{e^{-(\|f_2\|_\infty+\eta)\|\mu\|^2_{L^2_{T,p}}}}{1+\xi^2} & if\ \sigma\ is\ unbounded. \end{cases}$$

To compute the Ψ-CVaR$_\alpha$, we are lead to devise the usual *companion* procedure defined by

$$C_n = C_{n-1} - \gamma_n H_2\left(\xi_{n-1}, C_{n-1}, \mu_{n-1}, X^{(\mu_{n-1})}, W^{(n)}\right),\ C_0 \in L^1(\mathbb{P}), \tag{3.6}$$

where $H_2(\xi,C,\mu,\zeta,W) = C - w(\xi,\mu,\zeta,W)$ with

$$w(\xi,\mu,\zeta,W) = \xi + \frac{1}{1-\alpha}\left(\Psi(\varphi(\zeta))-\xi\right)\mathbf{1}_{\{\varphi(\zeta)>\xi\}} e^{-\int_0^T \langle f_2(\zeta)\mu_s, dW_s\rangle - \frac{1}{2}||f_2(\zeta)\mu_\cdot||^2_{L^2_{T,q}}}.$$

Consequently, the global procedure can be written in a more synthetic way

$$Z_n = Z_{n-1} - \gamma_n H\left(Z_{n-1}, W^{(n)}\right),\ Z_0 = (\xi_0, C_0, \theta_0, \mu_0), \tag{3.7}$$

with

$$H(z,W) := \big(H_1\left(\xi,\theta,X^{(\theta)},W\right), H_2\left(\xi,C,\mu,X^{(\mu)},W\right), H_3\left(\theta,\xi,X^{(-\theta)},W\right), H_{4,\lambda,\eta}\left(\mu,\xi,X^{(-\mu)},W\right)\big).$$

Next result shows that this procedure *a.s.* converges.

Theorem 3.3.3. *Suppose that* $(A2)_1$, $(\mathcal{H}_{b,\sigma})$, $(G_{\Psi,\varphi,\lambda})$ *and* $(A3)$ *are satisfied. Then, if the step sequence* $\gamma=(\gamma_n)_{n\geq 1}$ *satisfies* $(A1)$, *the recursive algorithm* (3.7) *satisfies*

$$Z_n \xrightarrow{a.s.} z^*_\alpha := \left(\xi^*_\alpha, C^*_\alpha, \theta^*_{\alpha,E}, \mu^*_{\alpha,E}\right), \quad as \quad n\to+\infty,$$

where ξ^*_α *is an* $\arg\min V$*-valued (squared integrable) random variable,* $C^*_\alpha := \Psi$-CVaR$_\alpha$, $\theta^*_{\alpha,E}$ *is an* $Argmin_{|E}\ Q_1(,\xi^*_\alpha)$*-valued (squared integrable) random variable and* $\mu^*_{\alpha,E}$ *is an* $Argmin_{|E}\ Q_2(,\xi^*_\alpha)$*-valued (squared integrable) random variable.*

Proof. Like for the static case, we firstly demonstrate that the tuple $t_n=(\xi_n,\theta_n,\mu_n)$ converges *a.s.* toward $t^*_\alpha := (\xi^*_\alpha,\theta^*_{\alpha,E},\mu^*_{\alpha,E}) \in \mathcal{T}^*$. Then, it will be straightforward that $(C_n)_{n\geq 1}$ *a.s.* converges toward C^*_α. The mean function of the three components algorithm is defined by

$$l(t) := \left(\mathbb{E}\left[H_1\left(\xi,\theta,X^{(\theta)},W\right)\right], \mathbb{E}\left[H_3\left(\theta,\xi,X^{(-\theta)},W\right)\right], \mathbb{E}\left[H_{4,\lambda,\eta}\left(\mu,\xi,X^{(-\mu)},W\right)\right]\right),$$

so that for every $t := (\xi, \theta, \mu) \in \mathbb{R} \times E \times E$,

$$\begin{aligned}\langle l(t), t - t^*_\alpha \rangle &= e^{-\frac{1}{2}||f_1||_\infty ||\theta||^2_{L^2_{T,q}}} \left(1 - \frac{1}{1-\alpha}\mathbb{P}(\varphi(X) > \xi)\right)(\xi - \xi^*_\alpha) \\ &\quad + \left\langle DQ_1(\theta), \theta - \theta^*_{\alpha,E} \right\rangle_E + C_{\lambda,\eta}(\mu) \left\langle DQ_2(\mu), \mu - \mu^*_{\alpha,E} \right\rangle_E > 0.\end{aligned}$$

It remains to check that

$$\left\|H_1(\xi, \theta, X^{(\theta)}, W)\right\|_2^2 + \left\|H_3\left(\theta, \xi, X^{(-\theta)}, W\right)\right\|_2^2 + \left\|H_{4,\lambda,\eta}\left(\mu, \xi, X^{(-\mu)}, W\right)\right\|_2^2 \leq C(1 + \|t\|^2) \tag{3.8}$$

in order to apply the Extended Robbins-Monro Theorem. A straightforward application of the proof of Theorem 4 in [62] shows that $\left\|H_3\left(\theta, \xi, X^{(-\theta)}, W\right)\right\|_2 \leq C(1 + \|\theta\|_{L^2_{T,p}})$ and $\left\|H_{4,\lambda,\eta}\left(\mu, \xi, X^{(-\mu)}, W\right)\right\|_2 \leq C(1 + \|\mu\|_{L^2_{T,p}})$ and we have

$$\begin{aligned}\mathbb{E}\left[H_1^2(\xi, \theta, X^{(\theta)}, W)\right] &\leq C\left(1 + e^{-||f_1||_\infty ||\theta||^2_{L^2_{T,q}}}\right. \\ &\quad \left. \times \mathbb{E}\left[\mathbf{1}_{\left\{\varphi(X^{(\theta)}) > \xi\right\}} e^{-2\int_0^T \left\langle f_1(X^{(\theta),s})\theta_s, dW_s\right\rangle - ||f_1(X^{(\theta),s})\theta_\cdot||^2_{L^2_{T,q}}}\right]\right), \\ &= C\left(1 + e^{-||f_1||_\infty ||\theta||^2_{L^2_{T,q}}}\right. \\ &\quad \left. \times \mathbb{E}\left[\mathbf{1}_{\{\varphi(X) > \xi\}} e^{-\int_0^T \left\langle f_1(X^{(\theta),s})\theta_s, dW_s\right\rangle + \frac{1}{2}||f_1(X^{(\theta),s})\theta_\cdot||^2_{L^2_{T,q}}}\right]\right) \\ &\leq C\left(1 + \mathbb{E}\left[e^{-\int_0^T \left\langle f_1(X^{(\theta),s})\theta_s, dW_s\right\rangle - \frac{1}{2}||f_1(X^{(\theta),s})\theta_\cdot||^2_{L^2_{T,q}}}\right]\right) \leq 2C,\end{aligned}$$

so that the linear growth assumption (3.8) is satisfied. This completes the proof. □

Like for the static framework, to circumvent the problem induced by the *a priori* optimal choice of the step sequence $\gamma_n = \frac{c}{n}$, $n \geq 1$ (the choice of c is more or less "blind" since it depends on the density of $\varphi(X)$ at ξ^*_α which is unknown), we are led to introduce the Cesàro mean $\bar{Z}_n$ of the algorithm (3.7) implemented with a slowly decreasing step $\gamma_n = \frac{c}{n^\beta}$ with $\frac{1}{2} < \beta < 1$ "a la Ruppert & Polyak". Next result shows that $\bar{Z}_n$ satisfies a Gaussian CLT with the optimal variance and the optimal rate $\sqrt{n}$.

Theorem 3.3.4. *Suppose that $(A2)_a$ with some $a > 1$, $(\mathcal{H}_{b,\sigma})$, $(G_{\Psi,\varphi,\lambda})$ and (A3) are satisfied and that the step sequence is $\gamma_n = \frac{c}{n^p}$ with $\frac{1}{2} < p < 1$ and $c > 0$. Then $\bar{Z}_n$ satisfies the Gaussian CLT*

$$\sqrt{n}\left(\bar{Z}_n - z^*_\alpha\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^*) \quad n \to +\infty \tag{3.9}$$

where Σ^* satisfies

$$\left\{\begin{array}{l} \Sigma^*_{1,1} = \frac{1}{f^2_{\varphi(X)}(\xi^*_\alpha)} \mathrm{Var}\left(\mathbf{1}_{\left\{\varphi(X^{(\theta^*_{\alpha,E})}) \geq \xi^*_\alpha\right\}} \right. \\ \qquad\qquad \left. \times e^{-\int_0^T \left\langle f_1(X^{(\theta^*_{\alpha,E}),s})\theta^*_{\alpha,E,s}, dW_s\right\rangle + \frac{1}{2}||f_1(X^{(\theta^*_{\alpha,E}),s})\theta^*_{\alpha,E,\cdot}||^2_{L^2_{T,q}}}\right) \\ \Sigma^*_{2,2} = \frac{1}{(1-\alpha)^2} \mathrm{Var}\left(\left(\Psi(\varphi(X^{(\mu^*_{\alpha,E})}) - \xi^*_\alpha\right) \mathbf{1}_{\left\{\varphi(X^{(\mu^*_{\alpha,E})}) \geq \xi^*\right\}} \right. \\ \qquad\qquad \left. \times e^{-\int_0^T \left\langle f_2(X^{(\mu^*_{\alpha,E}),s})\mu^*_{\alpha,E,s}, dW_s\right\rangle + \frac{1}{2}||f_2(X^{(\mu^*_{\alpha,E}),\cdot})\mu^*_{\alpha,E,\cdot}||^2_{L^2_{T,q}}}\right). \end{array}\right.$$

Proof. The proof follows the line of the proof of the equivalent result for the static framework. It consists in applying the Ruppert & Polyak's averaging principle (see Theorem 3.2 in [6]). □

Remark 2. • *Only a rough estimate of θ_α^* and μ_α^* is needed to reduce the asymptotic variance of the VaR-CVaR algorithm. That's why we don't specify the asymptotic variance of the I.S. procedure.*

• *For numerical implementation, as proposed in the static framework, we make the confidence level slowly move from $\alpha_0 = \frac{1}{2}$ to the final level α by introducing a deterministic confidence level sequence $(\alpha_n)_{n\geq 0}$. As demonstrated in [6], this leads to a new I.S. procedure $(\hat{\theta}_n, \hat{\mu}_n)_{n\geq 0}$ different from (3.4) that numericaly speeds up the initialization phase and improves the variance reduction. Finally, our optimal algorithm to compute the VaR_α and the Ψ-CVaR_α is given by the empirical mean $(\bar{\xi}_n, \bar{C}_n)$ where ξ_n is given by (3.5) and C_n is given by (3.6) with $(\hat{\theta}_n, \hat{\mu}_n)_{n\geq 0}$ as adaptive variance reducers.*

• *Finally as proposed in the finite dimensional setting, in practical implementation, we divide our procedure into two phases. Phase I is devoted to the estimation of the variance reducers $(\theta_{\alpha,E}^*, \mu_{\alpha,E}^*)$ using M iterations. Phase II produces some estimates of (ξ^*, C^*) based on (3.5) and (3.6) and its Cesàro mean with N iterations. During this phase, one can either keep on updating the I.S. parameters adaptively or "freeze" them at $(\hat{\theta}_M, \hat{\mu}_M)$ and only update (ξ_n, C_n). Note that since we only need a rough estimate of the optimal I.S. parameters, we set $M << N$. For more details about the practical implementation of the algorithm, we refer to Chapter 2 Section 3.*

3.4 Numerical examples

For the sake of simplicity, we focus on the computation of the VaR_α and CVaR_α ($\Psi \equiv Id$). We also consider three values of the confidence level $\alpha = 95\%,\ 99\%,\ 99.5\%$.

3.4.1 Asian option

We consider a process $(X_t)_{t\in[0,T]}$ solution of the following SDE:

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma\left(X_t\right)\mathrm{d}W_t,\ X_0 = x \in \mathbb{R}.$$

The portfolio is composed of a short position on European Asian Call option. Such options are of particular interest and importance to prevent price manipulations and represent a large proportion of the options traded in commodity markets. For instance, most options on oil are Asian options. The value of the loss of the portfolio at maturity T is defined by

$$\varphi(X) = \left(A(T) - K\right)_+ - e^{rT}C_0,$$

where C_0 is the price of the Asian option, r is the interest rate (and $r = 5\%$) and $A(T)$ is the average of the stock S_t over the time interval $[0, T]$, namely

$$A(T) = \frac{1}{T}\int_0^T S(u)\ \mathrm{d}u. \tag{3.10}$$

The problem of computing the price C_0 is known to be hard. Here, for the sake of simplicity we simply use a standard Monte Carlo method under a Black&Scholes model with a variance reduction technique based on the method developed by Kemna and Vorst [51]. It consists in using the random variable $e^{-rT}\left(e^Z - K\right)_+$ where $Z = \frac{1}{T}\int_0^T \log(S_u)\, du$ as a control variable. In order to approximate the integral $A(T)$, we use one of the schemes developed in [59]. To be precise, if we consider the continuous Euler scheme $\bar{X}^c$ obtained by extrapolation of the Brownian between two instants of discretization, *i.e.* for every $t \in [t_k, t_{k+1}]$,

$$\bar{X}^c_t = \bar{X}^c_{t_k} + b\left(\bar{X}^c_{t_k}\right)(t - t_k) + \sigma\left(\bar{X}^c_{t_k}\right)(W_t - W_{t_k}),\ \bar{X}^c_0 = x \in \mathbb{R}.$$

The step size will be noted $h = \frac{T}{N}$ and we define the times $t_k = k\frac{T}{N} = kh$. We set $T = 4\ months$ and $N = 100$. Now we can approximate the integral $A(T)$ by

$$\frac{1}{T}\sum_{k=0}^{N-1} \bar{X}^c_{t_k} h + \int_{t_k}^{t_{k+1}} \left(\bar{X}^c_s - \bar{X}^c_{t_k}\right) \mathrm{d}s. \tag{3.11}$$

It is shown in [59] that the weak convergence of this scheme holds at the rate $n^{-\frac{3}{2}}$. To compute this scheme, at each time step, we have to simulate $W_{t_{k+1}}$ knowing W_{t_k} and $\left(\int_{t_k}^{t_{k+1}} W_s\ ds|\ W_{t_k}, W_{t_{k+1}}\right)$ using the fact that for every $s \in [t_k, t_{k+1}]$

$$\mathcal{L}\left(\int_{t_k}^{t_{k+1}} W_s\ ds|\ W_{t_k} = u, W_{t_{k+1}} = v\right) = \mathcal{N}\left(\frac{t_{k+1} - s}{h}u + \frac{s - t_{k+1}}{h}v, \frac{(t_{k+1} - s)(s - t_k)}{h}\right).$$

Under the Black&Scholes model, by using a Taylor expansion, one can approximate the integral by the scheme

$$\frac{1}{T}\sum_{k=0}^{N-1} X_{t_k}\left(h + \frac{rh^2}{2} + \sigma\int_{t_k}^{t_{k+1}} (W_u - W_{t_k})\,\mathrm{d}u\right). \tag{3.12}$$

We consider three different basis of $L^2([0,1], \mathbb{R})$

- a polynomial basis composed of the shifted Legendre polynomials $(\tilde{P}_n)_{n\geq 0}$ defined by

$$\forall n \geq 0,\ \forall t \in [0,1],\quad \tilde{P}_n(t) = P_n(2t-1)\ \text{ where }\ P_n(t) = \frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\left(\left(t^2 - 1\right)\right). \tag{ShLeg}$$

- the Karhunen-Loeve basis defined by

$$\forall n \geq 0,\ \forall t \in [0,1],\quad e_n(t) = \sqrt{2}\ \sin\left(\left(n + \frac{1}{2}\right)\pi t\right). \tag{KL}$$

- the Haar basis which is defined by

$$\forall n \geq 0,\ \forall k = 0, ..., 2^n - 1,\ \forall t \in [0,1],\ \psi_{n,k}(t) = 2^{\frac{k}{2}}\psi(2^k t - n), \tag{Haar}$$

where,

$$\psi(t) = \begin{cases} 1 & \text{if } t \in [0, \frac{1}{2}) \\ -1 & \text{if } t \in [\frac{1}{2}, 1) \\ 0 & \text{otherwise.} \end{cases}$$

Black&Scholes Model
First, we consider that X follows the classical Black&Scholes model with a volatility $\sigma = 50\%$. The strike of the call is set at $K = 115$ and the initial price at $x = 100$. Note that in this case the price C_0 of the asian call option is computed with the help of our control variate which can be written

$$e^{-rT}\left(xe^{\left(r-\frac{\sigma^2}{2}\right)\frac{T}{2}+\frac{\sigma}{T}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}W_u du}-K\right)_+.$$

The price of this call is approximately equal to 2.24. We experiment our algorithm with trivial drivers $f_1 \equiv f_2 \equiv 1$. For different basis, the results of our algorithm are summarized in the following table. The optimal I.S. parameters are obtained using 100 000 iterations in the Robbins-Monro procedure then, with this optimized parameters, we compute the VaR_α, the CVaR_α and the two variance ratios $\mathrm{RV}_{\mathrm{VaR}}$, $\mathrm{RV}_{\mathrm{CVaR}}$ with 500 000 trials. For the three different basis, the results are summarized in the three following tables. In figure 1 are depicted the optimal variance reducer θ^* (the I.S. parameter for the VaR algorithm) when the optimization is carried out on E_m for several values of m (2, 4 and 8) in the different basis mentioned above and for a confidence level $\alpha = 95\%$. The results are summarized in Tables 3.1, 3.2, 3.3.

Basis	Dim.	α	VaR	CVaR	$\mathrm{RV}_{\mathrm{VaR}}$	$\mathrm{RV}_{\mathrm{CVaR}}$
Constant	1	95%	14.1	24.47	3.8	9.03
		99%	30.84	40.48	8.11	20.21
		99.5%	37.71	47.10	11.95	25.16
ShLeg	2	95%	14.3	24.76	9.8	33.1
		99%	31.12	40.69	20.4	52.8
		99.5%	37.85	47.32	27.19	57.27
ShLeg	4	95%	14.26	24.72	9.5	34.24
		99%	31.05	40.65	21.3	47.6
		99.5%	37.78	47.33	23.23	55.54
ShLeg	8	95%	14.47	24.84	9.9	33.4
		99%	31.37	40.85	22.1	45.7
		99.5%	37.99	47.3	21.94	69.85

Table 3.1: VaR, CVaR and variance ratios obtained using the Legendre basis in the Black&Scholes model.

Local Volatility Model
Now, we consider the same product in a local volatility model (inspired by the CEV model) defined by

$$\mathrm{d}X_t = rX_t\mathrm{d}t + \sigma X_t \frac{X_t^\beta}{\sqrt{1+X_t^2}}\,\mathrm{d}W_t,\ X_0 = x,$$

with $r = 0.05$, $\sigma = 7$, $x = 100$ and $\beta = 0.5$. The price of the Asian Call option given by a crude Monte Carlo with Brownian interpolation after 100 000 trials is approximately equal to 4.16. The results are summarized in Tables 3.4, 3.5, 3.6.

Basis	Dim.	α	VaR	CVaR	RV_{VaR}	RV_{CVaR}
KL	2	95%	14.1	24.66	3.5	7.3
		99%	31.0	40.90	7.04	15.48
		99.5%	37.73	47.18	9.05	19.87
KL	4	95%	14.1	24.64	4.9	11.66
		99%	30.95	40.6	10.64	23.45
		99.5%	37.73	47.18	13.72	24.5
KL	8	95%	14.17	24.65	6.0	11.8
		99%	31	40.57	14.53	24.0
		99.5%	37.76	47.23	19.56	42.87

Table 3.2: VaR, CVaR and variance ratios obtained using the Karhunen-Loève basis in the Black&Scholes model.

Basis	Dim.	α	VaR	CVaR	RV_{VaR}	RV_{CVaR}
Haar	2	95%	14.28	24.73	6.9	23.5
		99%	31.04	40.9	16.8	35.4
		99.5%	37.82	47.25	19.23	39.58
Haar	4	95%	14.41	24.78	8.16	25.54
		99%	31.17	40.79	19.78	42.8
		99.5%	37.95	47.32	23.07	49.84
Haar	8	95%	14.38	24.79	8.0	29.53
		99%	31.05	40.73	18.97	41.74
		99.5%	37.84	47.23	24	51.46

Table 3.3: VaR, CVaR and variance ratios obtained using the Haar basis in the Black&Scholes model.

3.4.2 Power plant

Like for the static framework, we consider a portfolio composed of a short position in a power plant that produces electricity day by day with a maturity of $T = 3\ months$ (we set N=100) and 100 long positions in calls on electricity day-ahead price with the same strike $K = 60$. We consider the Black&Scholes model, electricity's and gas's initial spot prices are $S_0^e = 40$ \$/MWh and $S_0^g = 3$ \$/MMBTU (BTU: British Thermal Unit) with a Heat Rate equals $h_R = 10 Btu/kWh$ and generation costs $C = 5/MWh$. The two brownian motions have a correlation of 0.4. The payoff can be written

$$\varphi(X) = \sum_{k=1}^{N} \left(e^{r(T-t_k)} \left(S_{t_k}^e - h_R S_{t_k}^g - C \right)_+ - P_0^c e^{rT} \right) + \left(e^{rT} C_0 - e^{r(T-t_k)} \left(S_{t_k}^e - K \right)_+ \right)$$

For different basis, the results are reported in the tables 3.7, 3.8, 3.9.

In the static framework, we also considered this example with different parameters and obtained the following results (see Table 3.10). For more details we refer to [6].

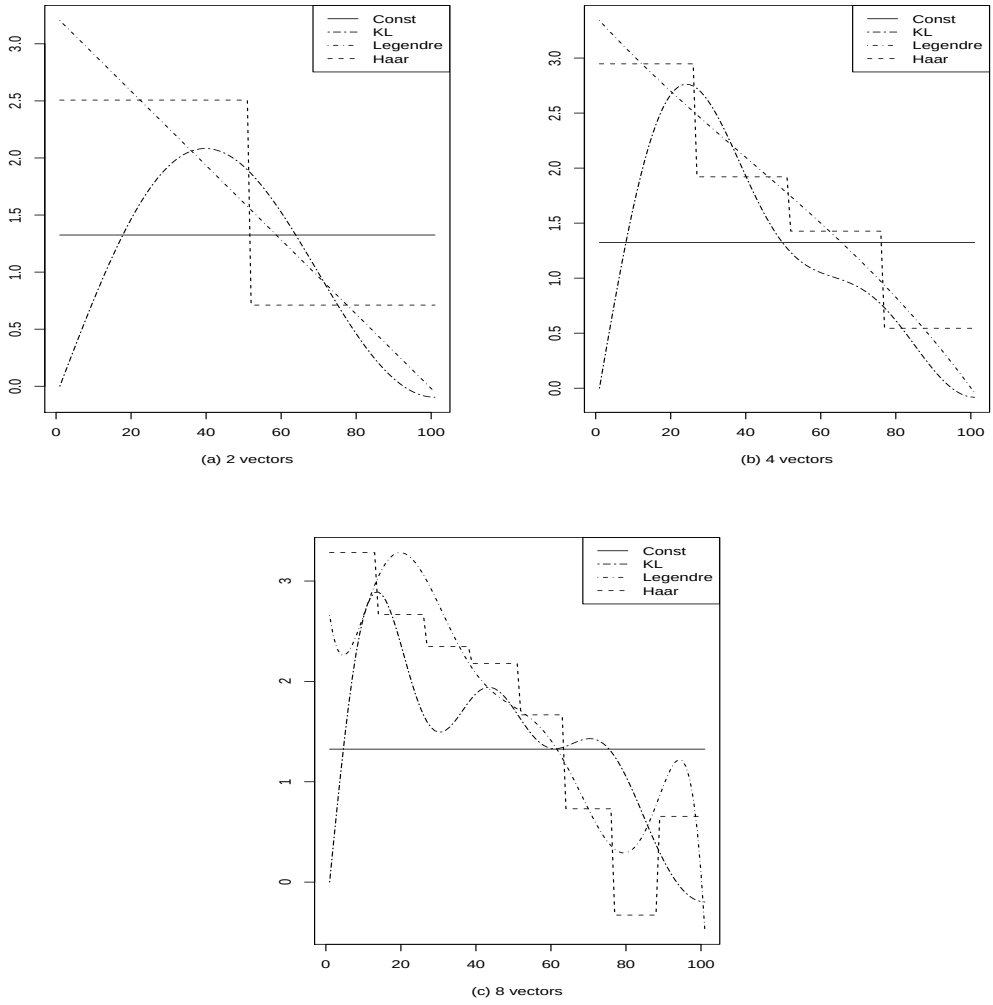


Figure 3.1: Optimal θ for different basis obtained with our algorithm using 100 000 steps and with a confidence level $\alpha = 95\%$.

3.4.3 Adaptive case: Up & In Put option

We consider a portfolio composed of short position on an Up & In Put option of strike K and barrier L. This option is activated when the underlying process X moves up and hits the barrier L. The value of the loss of the portfolio at maturity T is given by

$$\varphi(X) = (K - X_T)_+ \mathbf{1}_{\left\{\max_{0\leq t\leq T} X_t \geq L\right\}} - e^{rT} P_0$$

where P_0 is the price of the option. A standard approach to price the option is to consider the continuous Euler scheme $\bar{X}$ of step $t_k = k\frac{T}{N}$ obtained by extrapolation of the brownian bridge between two instants of discretization. For every $t \in [t_k, t_{k+1}]$, we can write

$$\bar{X}_t = \bar{X}_{t_k} + b(\bar{X}_{t_k})(t - t_k) + \sigma(\bar{X}_{t_k})(W_t - W_{t_k}),\ \bar{X}_0 = x_0 \in \mathbb{R}.$$

Basis	Dim.	α	VaR	CVaR	$\mathrm{RV}_{\mathrm{VaR}}$	$\mathrm{RV}_{\mathrm{CVaR}}$
Constant	1	95%	22.77	35.40	3.77	8.29
		99%	44.77	54.56	8.78	15.04
		99.5%	51.06	61.36	10.29	16.71
ShLeg	2	95%	23.12	35.60	9.12	22.70
		99%	43.23	54.10	15.5	23.4
		99.5%	51.36	61.53	25.20	36.86
ShLeg	4	95%	23.04	35.62	9.8	24.3
		99%	43.46	54.17	19.38	29.37
		99.5%	52.23	61.81	23.6	35.9
ShLeg	8	95%	23.26	35.71	11.2	23.7
		99%	43.65	54.28	20.5	33.60
		99.5%	51.44	61.65	23.81	37.35

Table 3.4: VaR, CVaR and variance ratios obtained using the Legendre basis in the local volatility model.

Basis	Dim.	α	VaR	CVaR	$\mathrm{RV}_{\mathrm{VaR}}$	$\mathrm{RV}_{\mathrm{CVaR}}$
KL	2	95%	22.85	35.52	3.33	6.95
		99%	43.34	54.17	6.55	12.08
		99.5%	51.10	61.54	8.89	17.03
KL	4	95%	22.95	35.57	4.79	11.60
		99%	43.26	54.13	10.81	17.56
		99.5%	51.11	61.48	14.96	23.03
KL	8	95%	22.92	35.54	5.80	13.00
		99%	43.33	54.19	14.16	22.70
		99.5%	51.13	61.53	18.72	29.24

Table 3.5: VaR, CVaR and variance ratios obtained using the Karhunen-Loève basis in the local volatility model.

By preconditioning,

$$\mathbb{E}\left[(K-\bar{X}_T)_+\mathbf{1}_{\left\{\max_{0\le t\le T}\bar{X}_t\ge L\right\}}\right]=\mathbb{E}\left[(K-\bar{X}_T)_+\left(1-\prod_{k=0}^{N-1}p(\bar{X}_{t_k},\bar{X}_{t_{k+1}})\right)\right],$$

where $p(x_k,x_{k+1})=\mathbb{P}\left(\max_{t_k\le t\le t_{k+1}}\bar{X}_t\le L\,\middle|\,\left(\bar{X}_{t_k},\bar{X}_{t_{k+1}}\right)=(x_k,x_{k+1})\right)$ is the probability of non exit of some brownian bridge. Using the law of the brownian bridge (see for example [42]), we can write

$$p(x_k,x_{k+1})=\begin{cases}1-e^{-\frac{2(L-x_k)(L-x_{k+1})}{(t_{k+1}-t_k)\sigma^2(x_k)}} & \text{if } L\ge\max(x_k,x_{k+1}),\\ 0 & \text{, otherwise.}\end{cases}\tag{3.13}$$

In the following simulation, we consider the classical Black&Scholes model. We set the number of steps of the Euler scheme N to 100, the interest rate r to 4%, the volatility to 70%. The strike of the option is set at $K=80$ and the barrier level at

Basis	Dim.	α	VaR	CVaR	$\mathrm{RV}_{\mathrm{VaR}}$	$\mathrm{RV}_{\mathrm{CVaR}}$
Haar	2	95%	23.06	35.63	6.88	16.86
		99%	43.39	54.23	17.11	28.47
		99.5%	51.18	61.49	21.43	31.22
Haar	4	95%	23.21	35.71	8.16	20.82
		99%	43.55	54.30	20.0	30.68
		99.5%	51.29	61.63	25.90	35.44
Haar	8	95%	23.19	35.68	8.3	22.41
		99%	43.49	54.25	21.89	34.91
		99.5%	51.26	61.58	26.29	38.53

Table 3.6: VaR, CVaR and variance ratios obtained using the Haar basis in the local volatility model.

Basis	Dim.	α	VaR	CVaR	$\mathrm{RV}_{\mathrm{VaR}}$	$\mathrm{RV}_{\mathrm{CVaR}}$
Constant	1	95%	386.81	1244.23	3.2	4.1
		99%	812.14	1468.08	14.40	53
		99.5%	980.28	1486.65	31.21	85.2
ShLeg	2	95%	390.29	1251.24	5.24	26.76
		99%	815.17	1472.38	28.89	67.2
		99.5%	981.14	1484.62	62.11	120.9
ShLeg	4	95%	389.82	1251.72	5.3	25.63
		99%	813.71	1471.61	28.15	66.54
		99.5%	982.13	1481.98	65.2	118.5
ShLeg	8	95%	385.97	1267.41	5.5	24.5
		99%	819.53	1473.15	29.35	75.2
		99.5%	978.64	1485.73	64.03	120.3

Table 3.7: VaR, CVaR and variance ratios obtained using the Legendre basis for the power plant portfolio in a Black&Scholes model.

$L = 110$. We set the price P_0 to the one given by a crude Monte Carlo after 500 000 trials. We obtain $P_0 = 1.73$ with a variance of 28.85.

We compare the efficiency of our algorithm with a trivial driver $f \equiv 1$ and with the non-trivial driver f defined for $t = t_k$ by

$$f(t, \zeta^t) = (\bar{p}_k \quad 1 - \bar{p}_k) \text{ with } \bar{p}_k = \prod_{i=0}^{k-1} p(\zeta^{t_i}, \zeta^{t_{i+1}}),$$

where p is defined by (3.13). There is no extra-computation for the drivers since the probabilities are already computed for the brownian bridge interpolation. In the adaptive case, for $t = t_k$, the optimal parameter can be written $\theta_{t_k} = \theta_1 \bar{p}_k + \theta_2(1-\bar{p}_k)$ with $(\theta_1, \theta_2) \in \mathbb{R}^2$. In the trivial-driver case, the optimal parameter is constant and can be written $\theta_{t_k} = \theta_1 \in \mathbb{R}$ for every $t = t_k$. The results for different confidence levels with the optimal parameter for the VaR I.S. procedure are reported in table 3.11.

Basis	Dim.	α	VaR	CVaR	RV_{VaR}	RV_{CVaR}
KL	2	95%	387.36	1248.19	3.4	5.6
		99%	818.90	1458.28	6.32	15.02
		99.5%	982.08	1485.43	9.56	27.44
KL	4	95%	388.27	1247.25	4.1	8.4
		99%	817.40	1465.55	8.9	23.30
		99.5%	977.07	1492.10	13.29	39.46
KL	8	95%	386.76	1252.09	4.3	11.2
		99%	819.29	1461.08	11.21	30.73
		99.5%	980.13	1489.30	16.32	49.45

Table 3.8: VaR, CVaR and variance ratios obtained using the Karhunen-Loève basis for the power plant portfolio in a Black&Scholes model.

Basis	Dim.	α	VaR	CVaR	RV_{VaR}	RV_{CVaR}
Haar	2	95%	389.67	1236.90	3.95	13.23
		99%	818.74	1457.53	12.33	36.15
		99.5%	981.03	1479.13	18.58	62.88
Haar	4	95%	392.97	1228.57	4.45	17.27
		99%	819.85	1452.08	14.7	43.6
		99.5%	988.29	1480.01	22.88	79.66
Haar	8	95%	384.54	1260.15	4.4	17.4
		99%	814.14	1477.63	14.49	36.95
		99.5%	983.27	1482.86	22.15	65.3

Table 3.9: VaR, CVaR and variance ratios obtained using the Haar basis for the power plant portfolio in a Black&Scholes model.

Number of steps	α	VaR	CVaR	RV_{VaR}	RV_{CVaR}
10 000	95%	115.7	150.5	3.4	6.8
	99%	169.4	196	8.4	12.9
	99.5%	186.3	213.2	13.5	20.3
100 000	95%	118.7	150.5	4.5	8.7
	99%	169.4	195.4	12.6	17.5
	99.5%	188.8	212.9	15.6	29.5
500 000	95%	119.2	150.4	5	9.2
	99%	169.8	195.7	13.1	18.6
	99.5%	188.7	212.8	17	29

Table 3.10: VaR, CVaR and variance ratios obtained using the finite dimensional I.S. algorithm for the power plant portfolio.

	α	VaR	CVaR	$\mathrm{RV}_{\mathrm{VaR}}$	$\mathrm{RV}_{\mathrm{CVaR}}$	θ_1^*	θ_2^* (adaptive case only)
Trivial	95%	13.8	22.4	1.3	1.7	-0.47	
driver	99%	27.5	33.0	1.7	2.3	-0.59	
	99.5%	31.8	36.6	1.8	2.5	-0.65	
Adaptive	95%	15.3	24.2	4.4	13.8	0.58	-4.12
case	99%	28.3	34.1	7.6	28.2	0.22	-3.6
	99.5%	32.5	37.7	9.1	54.0	0.03	-3.2

Table 3.11: VaR, CVaR and variance ratios obtained for the Up & In Put option in the Black&Scholes model.

Chapter 4

VaR-CVaR algorithm using quasi-stochastic approximation

A part of this work appeared in *Monte-Carlo and Quasi Monte-Carlo Methods 2008*.

Abstract: The aim of the paper is to extend the method of finding Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) by stochastic approximation, as introduced in [6], to quasi-random innovations. Using theoretical results about the discrepancy of Jordan measurable subsets of $[0,1]^q$ ($q \geq 1$), we establish a convergence theorem for the global Robbins-Monro algorithm under somewhat more restrictive assumptions than in the random case. Several simulations are provided to assert the superiority of this deterministic algorithm on the random one. In particular, we show that devising an I.S. algorithm in this context does accelerate the convergence of the VaR-CVaR algorithm.

Keywords: VaR, CVaR, Stochastic Approximation, Quasi-Monte Carlo, sequence with low discrepancy.

4.1 Introduction

Value-at-risk (VaR) and Conditional Value-at-risk (CVaR) are two widely used risk measures in the practice of financial risk management. CVaR is known to have better properties than VaR. Following the terminology introduced in the seminal paper [3] about risk measures, VaR does not satisfy the set of requirements of coherent risk measures whereas the CVaR does. For a given confidence level $\alpha \in (0,1)$ and portfolio loss distribution L, the VaR at level α (VaR_α is the lowest amount not exceeded by the loss with probability α (usually $\alpha \geq 95\%$). The Conditional Value-at-Risk at level α (CVaR_α) is the conditional expectation of the portfolio losses beyond the VaR_α level.

Several methods have been proposed to estimate VaR using historical or Monte Carlo simulation. The most commonly used method is the inversion of the empirical loss distribution function. Approximations of the portfolio loss distribution known as "Delta approximation" (resp. "Delta-Gamma approximation") relying on a linear (resp. quadratic expansion) are often used. For more details about those methods, we refer to [17], [39], [38].

In the context of portfolio optimization to reduce the CVaR, it is demonstrated in [77] that it is possible to approximate both VaR and CVaR (actually to optimize CVaR) by solving an approximation of a convex optimization problem which reduces to a linear programming problem. However, the linear programming problem is subject to huge number of linear constraints so that it remains difficult to handle in practice. In [6], the authors relies on a stochastic approximation algorithm to estimate simultaneously VaR and CVaR. This algorithm is derived from the fact that they are solutions of a convex optimization problem as demonstrated in [77]. It is shown that the convergence rate of this algorithm to its target satisfies a Gaussian Central Limit Theorem under standard assumptions. However, it is well known that in this kind of problem, we are interested by events that are observed with a very low probability (usually less that 5%). Thus, as a necessary improvement, a variance reduction technique based on an adaptive R.M. importance sampling was proposed and combined with the original VaR-CVaR algorithm (see [6] for more details). It has been shown in [39] that I.S. based on a large deviation analysis of a rare event can lead to efficient variance reduction. However, as noticed in [41], importance sampling estimators based on a large deviation analysis can have variance that increases with the rarity of the event and even infinite variance. In [25], a quantile estimator based on the inversion of the empirical weighted function with a projected version (on convex compact set) of a Robbins-Monro I.S. algorithm is introduced to find the optimal change of measure. The procedure proposed in [6] to select the optimal I.S. parameters is an unconstrained adaptive Robbins-Monro algorithm which somehow follows the new approach introduced in [62].

In this paper, we investigate the VaR-CVaR algorithm without I.S. in the case of quasi-random innovations, *i.e.* the case where the innovations are uniformly distributed on $[0,1]^q$ ($q \geq 1$) and have low discrepancy. Given the performance of quasi-Monte Carlo method for numerical integration, it is a rather natural idea to try plugging quasi-random numbers into a recursive stochastic approximation procedure instead of usual pseudo-random numbers. In the seminal paper [58], it is shown that it may significantly accelerate the convergence of the procedure like it does in numerical integration when using sequences with low discrepancy instead

of Monte-Carlo simulations. From a theoretical point of view, the two main results are based on a contraction assumption and on a stringent boundedness assumption. In the first framework, which is purely one-dimensional, some a priori error bounds emphasize that quasi-stochastic approximation does accelerate the convergence rate of the procedure.

Including several other frameworks than quasi-random innovations, quasi-stochastic approximation has been revisited recently in [60] generalising results obtained in [58] to multi-dimensional setting without boundedness assumption. The main result is based on averaging assumption on innovations and more classical Lyapunov assumptions (mean reverting and linear growth assumptions). However, those results do not apply directly to the VaR-CVaR algorithm since in most cases the averaging assumption is not satisfied.

This paper presents a convergence result for the VaR-CVaR Robbins-Monro (R.M.) algorithm originally investigated in [6]. We use general results about Jordan discrepancy (see [70] and [71]) so that the averaging assumption is satisfied.

This paper is organized as follows: Section 2 is devoted to the VaR-CVaR algorithm and low discrepancy sequences where we recall some definitions and characterizations. In Section 3, we describe our framework and our assumptions which lead us to the statement and the proof of the main convergence theorem. Section 4 gives numerical results. In particular we show that using low discrepancy sequences instead of pseudo-random numbers may accelerate significantly the convergence of the original procedure.

4.2 Some background: sequences with low discrepancy and VaR-CVaR algorithm

The purpose of this first part is to recall some classical results about sequences with low discrepancy in order to justify their use instead of i.i.d. pseudo-random numbers. Then, we will recall the R.M. algorithm proposed in [6] to estimate the VaR_α and the CVaR_α of a portfolio loss distribution L.

4.2.1 Sequences with low discrepancy: Definitions and characterizations

Definition 4.2.1. A $[0,1]^q$-valued sequence $(u_n)_{n\geq 1}$ is uniformly distributed (u.d.) on $[0,1]^q$ if

$$\frac{1}{n}\sum_{k=1}^{n}\delta_{u_k} \overset{(\mathbb{R}^q)}{\Longrightarrow} \mathcal{U}\left([0,1]^q\right), \quad n \to +\infty,$$

where for any subset A of $\mathbb{R}^q$, δ_u denotes the unit mass at u defined by $\delta_u(A) = \mathbf{1}_A(u)$, $\overset{(\mathbb{R}^q)}{\Longrightarrow}$ stands for the weak convergence of probability measures on $(\mathbb{R}^q, \mathcal{B}or(\mathbb{R}^q))$ and $\mathcal{U}\left([0,1]^q\right)$ denotes the uniform distribution on $[0,1]^q$.

The following proposition is a classical result in the theory of uniform distribution of sequences (see e.g. [56]). It introduces the notion of star discrepancy which can be viewed as a quantification of the definition of uniformly distributed sequences.

Proposition 4.2.1. *The $[0,1]^q$-valued sequence $(u_n)_{n\geq 1}$ is uniformly distributed on $[0,1]^q$ if and only if*

$$D_n^*(u) := \sup_{x\in[0,1]^q} \left| \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{[0,x]}(u_k) - \prod_{i=1}^{q} u_i \right| \to 0, \quad n \to +\infty.$$

$D_n^*(u)$ is called the star discrepancy.

We now introduce the notion of discrepancy for Jordan measurable subsets of $[0,1]^q$ (*i.e.* subsets with smooth boundary for which the characteristic function is Riemann integrable). This discrepancy will be usefull later when dealing with the VaR-CVaR procedure with quasi-random innovations. We refer to [71] for a proof.

Proposition 4.2.2. *Let $B \subseteq [0,1]^q$ and $\epsilon > 0$. Define*

$$B_\epsilon = \{u \in [0,1]^q : \ d(u,v) < \epsilon \ \text{ for some } v \in B\},$$
$$B_{-\epsilon} = \{u \in [0,1]^q : \ d(u,v) \geq \epsilon \ \text{ for all } v \in [0,1]^q \backslash B\},$$

where d denotes the standard Euclidean metric in $\mathbb{R}^q$. Let $\mathcal{M}_b$ be the family of all Lebesgue measurable $B \subseteq [0,1]^q$ for which

$$\lambda_q(B_\epsilon \backslash B) \leq b(\epsilon) \quad \text{and} \quad \lambda_q(B \backslash B_{-\epsilon}) \leq b(\epsilon),$$

for all $\epsilon > 0$, where λ_q denotes the Lebesgue measure on $(\mathbb{R}^q, \mathcal{B}or(\mathbb{R}^q))$, b is a positive non decreasing function satisfying $\lim_{\epsilon\to 0^+} b(\epsilon) = 0$. If $(u_n)_{n\geq 1}$ is uniformly distributed on $[0,1]^q$ then

$$D_n(\mathcal{M}_b, u) := \sup_{B\in\mathcal{M}_b} \left| \frac{1}{n}\sum_{k=1}^{n} \delta_B(u_k) - \lambda_q(B) \right| \to 0, \quad n \to +\infty.$$

Moreover, if the function b is of the form $b(\epsilon) = C\epsilon$, for all $\epsilon > 0$ and for some constant $C > 0$ then there exists $K > 0$ such that

$$D_n(\mathcal{M}_b, u) \leq K D_n^*(u)^{\frac{1}{q}}.$$

Actually, it has become customary to speak of low-discrepancy sequences if the sequence u satisfies $D_n^*(u) = O\left(N^{-1}(\log N)^q\right)$.

We now briefly discuss two important results for numerical integration using quasi-random innovations. It provides error bounds for the empirical $\frac{1}{n}\sum_{k=1}^{n} f(u_k)$ as estimator of $\mathbb{E}[f(U)]$ where $U \sim \mathcal{U}([0,1]^q)$ and f is a function with finite variation (in the measure sense or in the Hardy and Krause sense).

Proposition 4.2.3. *Let $u = (u_k)_{1\leq k\leq n}$ be a $[0,1]^q$-valued sequence and let f be a function with finite variation $V(f)$. Then*

- *(Koksma-Hlwaka Inequality)*

$$\left| \frac{1}{n}\sum_{k=1}^{n} f(u_k) - \int_{[0,1]^q} f(u)\lambda_q(du) \right| \leq V(f) D_n^*(u).$$

- *([19]) If $B \in \mathcal{M}_b$, we have*

$$\left| \frac{1}{n} \sum_{k=1,\ u_k\in B}^{n} f(u_k) - \int_B f(u)\lambda_q(du) \right| \leq \left(V(f) + f(1,\cdots,1)\right) D_n(\mathcal{M}_b, u).$$

Another usefull error bound is based on a Lipschitz assumption of f. The following theorem is due to Proinov (see [75])

Theorem 4.2.4. *Assume $\mathbb{R}^q$ is equipped with the ℓ^∞-norm ($|x|_\infty := \max_{1\leq i\leq q}|x_i|,\ x\in\mathbb{R}^q$). Let u be a u.d. sequence on $[0,1]^q$. Let $p_1,\cdots,p_n$ be a sequence of non negative numbers satisfying*

$$\sum_{k=1}^{n} p_k = 1.$$

Then, for every continuous function $f:[0,1]^q\to\mathbb{R}$

$$\left|\sum_{k=1}^{n} p_k f(u_k) - \int_{[0,1]^q} f(x)\lambda_q(dx)\right| \leq C_q w_f\left(D_n^*(u)^{\frac{1}{q}}\right),$$

where

$$w_f(\delta) := \sup_{x,y\in[0,1]^q,|x-y|_\infty\leq\delta} |f(x)-f(y)|,\quad \delta\in(0,1),$$

and $C_q\in(0,\infty)$ is a universal constant only depending on q. If $q=1$, $C_q=1$ and if $q\geq 2$, $C_q\in[1,4]$.

4.2.2 VaR-CVaR stochastic approximation algorithm

In this section we briefly recall the main tools and assumptions used to devise the VaR-CVaR algorithm investigated in [6].

We consider that the loss of the portfolio over the considered time horizon can be written as a function of a structural finite dimensional random vector, i.e., $L=\varphi(X)$ where X is a $\mathbb{R}^d$-valued random vector defined on the probability space $(\Omega,\mathcal{A},\mathbb{P})$ and $\varphi:\mathbb{R}^d\to\mathbb{R}$ is a Borel function. Thus, φ is the function describing the composition of the portfolio which remains fixed and X is a structural d-dimensional random vector used to model the market prices over a given time interval. We only rely on the fact it is possible to sample from the distribution of X. For instance, in a Black-Scholes framework, X is generally a vector of Brownian increments related to the Euler scheme of a diffusion. The VaR at level $\alpha\in(0,1)$ is the lowest α-quantile of the distribution $\varphi(X)$:

$$\mathrm{VaR}_\alpha(\varphi(X)) := \inf\left\{\xi \mid \mathbb{P}\left(\varphi(X)\leq\xi\right)\geq\alpha\right\}.$$

We assume that the distribution function of $\varphi(X)$ is continuous (i.e., without atoms) thus the $\mathrm{VaR}_\alpha := \xi_\alpha^*$ is the lowest solution of the equation:

$$\mathbb{P}(\varphi(X)\leq\xi_\alpha^*) = \alpha.$$

If the distribution function of $\varphi(X)$ is (strictly) increasing then the solution of the above equation is unique. Assuming that $\varphi(X)\in L^1(\mathbb{P})$, the $\mathrm{CVaR}_\alpha := C_\alpha^*$ is defined by:

$$\mathrm{CVaR}_\alpha(\varphi(X)) := \mathbb{E}\left[\varphi(X)|\varphi(X)\geq\mathrm{VaR}_\alpha(\varphi(X))\right].$$

The idea to devise a stochastic approximation algorithm to estimate the couple $(\xi_\alpha^*, C_\alpha^*)$ comes from the fact that it appears as the solution and the value of a convex optimisation problem. (see [77] and [6] for more details)

Proposition 4.2.5. *Let V be the function defined on $\mathbb{R}$ by: $\xi \mapsto \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(\varphi(X)-\xi)_+\right]$. Suppose that the distribution function of $\varphi(X)$ is continuous. Then, the function V is convex, differentiable and the $\mathrm{VaR}_\alpha(\varphi(X))$ is any point of the set:*

$$\arg\min\ V = \{\xi \in \mathbb{R} \mid V'(\xi) = 0\} = \{\xi \mid \mathbb{P}(\varphi(X) \le \xi) = \alpha\}$$

where V' is the derivative defined of V. Moreover, for every $\xi \in \mathbb{R}$, $V'(\xi) = \mathbb{E}\left[H_1(\xi, X)\right]$ where,

$$H_1(\xi, x) := 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(x)\ge\xi\}}.$$

Furthermore, $\mathrm{CVaR}_\alpha(\varphi(X)) = \min_{\xi\in\mathbb{R}} V(\xi)$.

Since we are looking for ξ for which $\mathbb{E}\left[H_1(\xi, X)\right] = 0$, we implement a stochastic gradient descent derived from the Lyapunov function V to approximate ξ_α^*, i.e., we use the RM algorithm:

$$\xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), \quad n \ge 1 \tag{4.1}$$

where $(X_n)_{n\ge1}$ is an i.i.d. sequence of random variables with the same distribution as X, independent of ξ_0, with $\mathbb{E}\left[|\xi_0|\right] < \infty$ and $(\gamma_n)_{n\ge1}$ is a positive deterministic step sequence (decreasing to 0) satisfying

$$\sum_{n\ge1}\gamma_n = +\infty \quad \text{and} \quad \sum_{n\ge1}\gamma_n^2 < +\infty. \tag{A1}$$

In order to estimate C_α^*, [6] proposed to devise an averaging procedure of the above quantile search algorithm with the same step sequence $(\gamma_n)_{n\ge1}$, namely $C_0 = 0$ and for $n = 1,\ 2,\ ...,$

$$C_n = C_{n-1} - \gamma_n H_2(\xi_{n-1}, C_{n-1}, X_n). \tag{4.2}$$

where $H_2(\xi, c, x) := c - \xi - \frac{1}{1-\alpha}\left(\varphi(x)-\xi\right)_+$. The resulting algorithm reads for $n \ge 1$:

$$\begin{cases} \xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), & \xi_0 \in \mathbb{R} \\ C_n = C_{n-1} - \gamma_n H_2(\xi_{n-1}, C_{n-1}, X_n), & C_0 = 0. \end{cases} \tag{4.3}$$

As soon as the distribution of $\varphi(X)$ is continuous, $\varphi(X) \in L^2(\mathbb{P})$ and that the step sequence satisfies (A1), the sequence $(\xi_n, C_n)_{n\ge1}$ *a.s.* converges toward $(\xi_\alpha^*, C^*\alpha)_{n\ge1}$ where ξ_α^* is an $\mathrm{Arg\,min}\, V$-valued random variable and $C_\alpha^* = \mathrm{CVaR}_\alpha(\varphi(X))$ (we refer to [6] for a proof).

To achieve the best convergence rate, we are led to introduce the Ruppert and Polyak's averaging principle (see [45] and [81]). If we set $\gamma_n = cn^{-p}$, with $\frac{1}{2} < p < 1$, $c > 0$ in (4.3) and compute the Cesàro means of both components

$$\begin{cases} \overline{\xi}_n := \frac{1}{n}\sum_{k=1}^n \xi_k = \overline{\xi}_{n-1} - \frac{1}{n}(\overline{\xi}_{n-1} - \xi_n) \\ \overline{C}_n := \frac{1}{n}\sum_{k=1}^n C_k = \overline{C}_{n-1} - \frac{1}{n}(\overline{C}_{n-1} - C_n) \end{cases} \tag{4.4}$$

where (ξ_k, C_k), $k \ge 0$ is defined by (4.3) then, provided that

$$\mathbb{E}\left[|\varphi(X)|^{2a}\right] < +\infty \ \text{ for some } a > 1, \tag{4.5}$$

and that the distribution of $\varphi(X)$ has a positive probability density $f_{\varphi(X)}$ on its support, we obtain asymptotically efficient estimators which satisfy the Gaussian CLT:

$$\sqrt{n}\begin{pmatrix} \overline{\xi}_n - \xi_\alpha^* \\ \overline{C}_n - C_\alpha^* \end{pmatrix} \overset{\mathcal{L}}{\to} \mathcal{N}(0,\Sigma) \tag{4.6}$$

where the asymptotic covariance matrix Σ is given by

$$\Sigma = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f^2_{\varphi(X)}(\xi^*)} & \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[(\varphi(X)-\xi^*)_+\right] \\ \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[(\varphi(X)-\xi^*)_+\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\left(\varphi(X)-\xi^*)_+\right. \end{pmatrix}. \tag{4.7}$$

The bottleneck in using the above algorithm lies in its very slow convergence owing to the fact that $\mathbb{P}(\varphi(X) > \xi^*) = 1-\alpha$ is close to 0 in practical implementations, so that we observe few significant replications to update our estimates. Moreover, in the bank and energy sectors, practitioners usually deal with huge portfolio composed by hundreds or thousands of risk factors and options. The evaluation step of $\varphi(X)$ may require a lot of computational time. Consequently, to achieve accurate estimates of both VaR_α and CVaR_α with reasonable computational effort, the above algorithm (4.4) needs to be speeded up by an IS procedure to recenter simulation "where things do happen", i.e., scenarios for which $\varphi(X)$ exceeds ξ. In [6], an IS algorithm based on an adaptive RM procedure is combined with (4.4). However, in this deterministic framework, we no longer need to reduce the asymptotic variance.

4.3 VaR-CVaR algorithm using quasi-stochastic approximation

The aim of this section is to describe the framework of the VaR-CVaR algorithm using quasi-random innovations and to establish the convergence theorem of this deterministic procedure. We aim at building a framework in which the averaging assumption introduced in [60] is satisfied. As a consequence, we will use the main theorem of [60] in order to establish the convergence of the VaR-CVaR procedure in this deterministic context. Then, we will discuss the possibility of plugging a recursive I.S. R.M. algorithm in order to accelerate the convergence.

4.3.1 VaR-CVaR quasi-stochastic approximation algorithm

Generally, the d-dimensional random variable X can be simulated from the uniform distribution $\mathcal{U}([0,1]^q)$, with $q \geq d$ by standard methods like inverse distribution function, Box-Müller, etc, so that $X = \Psi(U)$, with $\Psi : [0,1]^q \to \mathbb{R}^d$. Instead of the pseudo-random variable U, we will use a low-discrepancy sequence u. We will need the following assumption on the function φ and Ψ:

Assumption 4.3.1. *The function $\varphi o \Psi : [0,1]^q \to \mathbb{R}$ is Lipschitz.*

We denote by F the loss distribution function, *i.e.* the distribution function of $\varphi(\Psi(U))$. From a theoretical point of view, the convergence of the procedure can be

derived from the weak convergence:

$$F_n^{\Psi} := \frac{1}{n}\sum_{k=1}^{n} \delta_{\Psi(u_k)} \overset{(\mathbb{R}^d)}{\Rightarrow} F.$$

We write $D_n^*(u,\Psi)$ for the star discrepancy of the first n terms of u associated to the system defined naturally by

$$D_n^*(u,\Psi) := \sup_{\xi\in\mathbb{R}} \left| \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\Psi(u_k)\leq\xi} - F(\xi) \right|. \tag{4.8}$$

Note that as soon as u is u.d., $D_n^*(u,\Psi) \to 0$, as $n \to +\infty$.

Proposition 4.3.2. *Suppose that assumption 4.3.1 is satisfied and that the distribution function of $L = \varphi(\Psi(U))$ is Lipschitz. Then, for all $\xi \in \mathbb{R}$*

$$(\varphi o\Psi)^{-1}\Big(\,]-\infty,\xi\,] \Big) \in \mathcal{M}_b,$$

where for all $\epsilon > 0$, $b(\epsilon) = C\epsilon$ for some constant $C > 0$. If u is a u.d. sequence with low discrepancy then,

$$l_n := \max_{1\leq k\leq n} kD_k^*(u,\Psi) = O\left(n^{1-\frac{1}{q}}\log(n)\right), \quad n \geq 1.$$

Proof. Let $\xi \in \mathbb{R}$ and $B = (\varphi o\Psi)^{-1}\Big(\,]-\infty,\xi\,] \Big)$. By definition of B_ϵ, we have

$$B_\epsilon \subset (\varphi o\Psi)^{-1}\Big(\,]-\infty,\xi + w(\varphi o\Psi,\epsilon)\,] \Big),$$

and,

$$B_{-\epsilon} \supset (\varphi o\Psi)^{-1}\Big(\,]-\infty,\xi - w(\varphi o\Psi,\epsilon)\,] \Big),$$

where

$$w(\varphi o\Psi) = \sup_{x,y\in[0,1]^q} \frac{|\varphi o\Psi(x) - \varphi o\Psi(y)|}{|x-y|_\infty},$$

and $|x|_\infty := \max_{1\leq i\leq q}|x_i|$.
Consequently,

$$\begin{aligned}\lambda_q\left(B_\epsilon \backslash B\right) &\leq \lambda_q\left((\varphi o\Psi)^{-1}\left(\,]-\infty,\xi + w(\varphi o\Psi)\epsilon\,]\right)\right),\\ &= F(\xi + w(\varphi o\Psi)\epsilon) - F(\xi),\\ &\leq C\epsilon,\end{aligned}$$

and, in the same way, we have

$$\lambda_q\left(B\backslash B_{-\epsilon}\right) \leq F(\xi) - F(\xi - w(\varphi o\Psi)\epsilon) \leq C\epsilon.$$

Proposition 4.2.2. implies that

$$D_n^*\left(u,\Psi\right) = O\left(n^{-\frac{1}{q}}\log(n)\right),$$

so that,

$$l_n = \max_{1\leq k\leq n} kD_k^*(u,\Psi) = O\left(n^{1-\frac{1}{q}}\log(n)\right), \quad n \geq 1.$$

This completes the proof. □

The above convergence rate obtained for the star discrepancy of the system corresponds to the averaging assumption introduced in [60]. Let u be u.d. sequence in $[0,1]^q$, the VaR-CVaR quasi stochastic algorithm is defined for $n \geq 1$ by

$$\xi_n = \xi_{n-1} - \gamma_n K_1(\xi_{n-1}, u_n) \tag{4.9}$$
$$C_n = C_{n-1} - \gamma_n K_2(\xi_{n-1}, C_{n-1}, u_n). \tag{4.10}$$

where for all $\xi \in \mathbb{R}$, $u \in [0,1]^q$, $K_1(\xi, u) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(\Psi(u))\geq\xi\}}$ and $K_2(\xi, c, u) := c - \xi - \frac{1}{1-\alpha}\left(\varphi(\Psi(u)) - \xi\right)_+$.
Now we are in position to state the main convergence theorem.

Theorem 4.3.3. *Suppose that assumption 4.3.1 is satisfied and that the distribution function of $L = \varphi(\Psi(U))$ is Lipschitz. Let u be u.d. sequence in $[0,1]^q$ with low discrepancy. Let $\gamma = (\gamma_n)_{n\geq 1}$ be a non-negative non-increasing sequence of gain parameters satisfying*

$$\sum_{n\geq 1}\gamma_n = +\infty, \quad \gamma_n l_n \longrightarrow 0, \quad \sum_{n\geq 1}\max\left(|\Delta\gamma_{n+1}|, \gamma_n^2\right) l_n < +\infty. \tag{4.11}$$

Then the recursive procedure defined by (4.9) *and* (4.10) *converge toward* $(\xi_\alpha^*, C_\alpha^*)$.

Proof. In order to proove the convergence of the sequence $(\xi_n)_{n\geq 1}$, we use Theorem 2.1 of [60]. Let L be the continuously differentiable function defined for all $\xi \in \mathbb{R}$ by $L(\xi) = \frac{1}{2}(\xi - \xi^*)^2$. This functions satisfies

$$\nabla L \text{ is Lipschitz continuous and } |\nabla L|^2 \leq C(1 + L).$$

Since K_1 is bounded, it satisfies the linear growth assumption which in this deterministic framework is

$$\forall\, \xi \in \mathbb{R},\ \forall\, u \in [0,1]^q, \quad |K_1(\xi, u)|^2 \leq C\left(1 + L(\xi)\right). \tag{4.12}$$

The local mean reverting assumption is satisfied since

$$\langle \nabla(\xi), K_1(\xi, u) - K_1(\xi_\alpha^*, u)\rangle = \frac{1}{1-\alpha}(\xi - \xi_\alpha^*)\left(\mathbf{1}_{\{\varphi(\Psi(u))\leq\xi\}} - \mathbf{1}_{\{\varphi(\Psi(u))\leq\xi_\alpha^*\}}\right) > 0,$$

for $\xi \neq \xi_\alpha^*$ and for all $u \in [0,1]^q$. Proposition 4.3.2 implies that

$$\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\Psi(u_k)\leq\xi_\alpha^*} - F(\xi_\alpha^*)\right| = O(n^{-\frac{1}{q}}\log(n)),$$

thus, owing to Theorem 2.1 in [60], the sequence $(\xi_n)_{n\geq 1}$ converges toward ξ_α^*, as $n \to +\infty$. In order to prove the convergence of $(C_n)_{n\geq 1}$ toward C_α^*, we set for convenience $\gamma_0 := 1 + \sup_{n\geq 1}\gamma_n$. Then, one defines recursively a sequence $(\Delta_n)_{n\geq 1}$ by

$$\Delta_{n+1} = \Delta_n \frac{\gamma_{n+1}}{\gamma_n}\frac{\gamma_0}{\gamma_0 - \gamma_{n+1}},\ n \geq 0,\ \Delta_0 = 1.$$

Elementary computations show by induction that

$$\gamma_n = \gamma_0\frac{\Delta_n}{S_n},\ n \geq 0,\ \text{with } S_n := \sum_{k=0}^{n}\Delta_k. \tag{4.13}$$

Furthermore, it follows from (4.13) that for every $n \geq 1$, $\log(S_n) \geq \frac{1}{\gamma_0}\sum_{k=1}^{n}\gamma_k$, which implies that $\lim_n S_n = +\infty$.
Now using (4.10) and (4.13), one gets for every $n \geq 1$

$$S_nC_n = S_{n-1}C_{n-1} + \Delta_n\left(\xi_{n-1} + \frac{1}{1-\alpha}(F(u_k) - \xi_{k-1})_+\right).$$

This implies that

$$C_n = \frac{1}{S_n}\left(\sum_{k=1}^{n}\Delta_k\xi_{k-1}\right) + \frac{1}{1-\alpha}\frac{1}{S_n}\left(\sum_{k=1}^{n}\Delta_k(F(u_k) - \xi_{k-1})_+\right),$$

so that

$$\begin{aligned} C_n - C_\alpha^* = \frac{1}{S_n}&\left(\sum_{k=1}^{n}\Delta_k(\xi_{k-1} - \xi_\alpha^*)\right) \\ &+ \frac{1}{1-\alpha}\frac{1}{S_n}\left(\sum_{k=1}^{n}\Delta_k\left((F(u_k) - \xi_{k-1})_+ - \mathbb{E}\left[(F(U) - \xi_\alpha^*)_+\right]\right)\right). \end{aligned} \tag{4.14}$$

The first term of the right-hand side of (4.14) converges toward 0 owing to Cesàro Lemma and the convergence of $(\xi_n)_{n\geq 1}$ toward ξ_α^*. The second term can be written

$$\begin{aligned} \frac{1}{S_n}\sum_{k=1}^{n}\Delta_k\left((F(u_k) - \xi_{k-1})_+ - \mathbb{E}\left[(F(U) - \xi_\alpha^*)_+\right]\right) = \\ \frac{1}{S_n}\sum_{k=1}^{n}\Delta_k\left((F(u_k) - \xi_{k-1})_+ - (F(u_k) - \xi_\alpha^*)_+\right) \\ + \frac{1}{S_n}\sum_{k=1}^{n}\Delta_k\left((F(u_k) - \xi_\alpha^*)_+ - \mathbb{E}\left[(F(U) - \xi_\alpha^*)_+\right]\right). \end{aligned}$$

For all $u \in [0,1]^q$, the functions $\xi \mapsto (F(u) - \xi)_+$ are Lipschitz so that

$$\frac{1}{S_n}\sum_{k=1}^{n}\Delta_k\left((F(u_k) - \xi_{k-1})_+ - (F(u_k) - \xi_\alpha^*)_+\right) \leq \frac{1}{S_n}\sum_{k=1}^{n}\Delta_k\left|\xi_{k-1} - \xi_\alpha^*\right| \to 0, \text{ as } n \to +\infty.$$

Now, Assumption 4.3.1 implies that the function $u \mapsto (F(u) - \xi_\alpha^*)_+$ is Lipschitz so that owing to Theorem 4.2.4

$$\left|\frac{1}{S_n}\sum_{k=1}^{n}\Delta_k(F(u_k) - \xi_\alpha^*)_+ - \mathbb{E}\left[(F(U) - \xi_\alpha^*)_+\right]\right| \leq C_q w(\varphi o\Psi)D_n^*(u)^{\frac{1}{q}} \to 0, \text{ as } n \to +\infty.$$

This completes the proof. □

Remark 3. *The choice of* $\gamma_n := \frac{c}{n^\rho}$ *with* $1 - \frac{1}{q} < \rho \leq 1$ *is always acceptable.*

4.3.2 Design of a faster procedure: I.S. quasi-stochastic algorithm

Obviously, there is no theoretical reason in this deterministic context to plug the I.S. procedure investigated in [62] and already used in [6] in a stochastic framework with the VaR-CVaR quasi-stochastic approximation algorithm. However, from a numerical point of view, we observe that it does accelerate the convergence of the original procedure as it did in the stochastic context.

Suppose that $F(X)$ is a square integrable random variable such that $\mathbb{P}(F(X) \neq 0) > 0$ and where X is a random vector with density function p over $\mathbb{R}^d$. The main idea of I.S. by translation, applied to the computation of $\mathbb{E}[F(X)]$, is to use the invariance of the Lebesgue measure by translation: it follows that for every $\theta \in \mathbb{R}^d$,

$$\mathbb{E}[F(X)] = \int_{\mathbb{R}^d} F(x)p(x)dx = \int_{\mathbb{R}^d} F(x+\theta)p(x+\theta)dx = \mathbb{E}\left[F(X+\theta)\frac{p(X+\theta)}{p(X)}\right]. \tag{4.15}$$

Among all these random vectors with the same expectation, we want to select the one with the lowest variance, *i.e.*, the one with lowest quadratic norm

$$Q(\theta) := \mathbb{E}\left[F^2(X+\theta)\frac{p^2(X+\theta)}{p^2(X)}\right], \quad \theta \in \mathbb{R}^d.$$

A reverse change of variable shows that:

$$Q(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right], \ \theta \in \mathbb{R}^d. \tag{4.16}$$

Under some assumptions on the probability density function p (see [62]), one shows that Q is finite, convex, differentiable and that $\lim_{|\theta|\to+\infty} Q(\theta) = +\infty$ so that $\arg\min Q = \left\{\theta \in \mathbb{R}^d \mid \nabla Q(\theta) = 0\right\} \neq \varnothing$. If ∇Q admits a representation as an expectation, then it is possible to devise a recursive R.M. procedure to approximate the optimal parameter θ^*, namely

$$\theta_n = \theta_{n-1} - \gamma_n L_1(\theta_{n-1}, X_n), \ n \geq 1 \tag{4.17}$$

where L_1 is naturally defined by the formal differentiation of Q, for every $x \in \mathbb{R}^d$:

$$L_1(\theta, x) = F^2(x)\frac{p(x)}{p^2(x-\theta)}\nabla p(x-\theta). \tag{4.18}$$

Since we have no knowledge about the regularity of F and do not wish to have any, we differentiate the second representation of Q in (4.16) and not the first one. However, the regular RM procedure (4.17) suffers from an instability issue coming from the fact that the classical sub-linear growth assumption in quadratic mean in the Robbins-Monro Theorem is only fulfilled when F is constant, due to the behaviour of the annoying term $p(x)/p(x-\theta)$ as θ goes to infinity.

Recently, IS using stochastic algorithm was deeply revisited in [62] to remove the constraints introduced by the original algorithm. Moreover, this construction is extended to a large class of probability distributions and to diffusion process. Thanks to another translation of the variable θ, it is possible to plug back the parameter θ

"into" $F(X)$, the function F having in common applications a known behaviour at infinity.

From now on, we assume that there exist two positive constants $a,\ C\ > 0$ such that

$$\forall x \in \mathbb{R}^d,\ |F(x)|^2 \leq C\ e^{a|x|}. \tag{4.19}$$

Now under (4.19), we can use another change of variable so that Q is differentiable on $\mathbb{R}^d$ with a gradient given by

$$\nabla Q(\theta) := \mathbb{E}\left[L_1(\theta, X)\right] = \mathbb{E}\left[F^2(X-\theta)\underbrace{\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}}_{W(\theta,X)}\right].$$

We make the following assumption on the probability density p of X

$$\exists b \in [1,2] \text{ such that } \begin{cases} (i) & \frac{|\nabla p(x)|}{p(x)} = O(|x|^{b-1}) \quad \text{as} \quad |x| \to \infty \\ (ii) & \exists\ \rho > 0 \text{ such that } \log\left(p(x)\right) + \rho|x|^b \text{ is convex.} \end{cases} \tag{4.20}$$

Now under condition (4.20), there exist here exist two constants A and B such that

$$|W(\theta, X)| \leq e^{2\rho|\theta|^b}(A|x|^{b-1} + A|\theta|^{b-1} + B) \tag{4.21}$$

so that this weight can always be controlled by a deterministic function of θ (for more details, one can refer to [62]). Then by setting,

$$\widetilde{W}(\theta, X) := e^{-2\rho|\theta|^b} W(\theta, X),$$

we can define L_2 by

$$L_2(\theta, x) := F^2(X-\theta)\widetilde{W}(\theta, x), \tag{4.22}$$

so that it satisfies the linear growth assumptions and

$$\left\{\theta \in \mathbb{R}^d \mid \mathbb{E}\left[L_2(\theta, X)\right] = 0\right\} = \left\{\theta \in \mathbb{R}^d \mid \nabla Q(\theta) = 0\right\},$$

and the sequence $(\theta_n)_{n\geq 1}$ defined by (4.17) with L_2 instead of L_1 satisfies $\theta_n \xrightarrow{a.s.} \theta^* \in \operatorname{Arg\,min} Q$.

This new version was the starting point to define two R.M. I.S. algorithms. Indeed, applied to the problem we are dealing with, the main idea is to twist (by translation) the distribution of X in order to minimize the asymptotic variance of the two components in the CLT (4.6): the asymptotic variances of the VaR_α and CVaR_α algorithm

$$\frac{\mathrm{Var}\left(\ \mathbf{1}_{\varphi(X)\geq\xi^*}\right)}{f^2_{\varphi(X)}(\xi^*)} = \frac{\alpha(1-\alpha)}{f^2_{\varphi(X)}(\xi^*)} \quad \text{and} \quad \frac{\mathrm{Var}\left((\varphi(X)-\xi^*)_+\right)}{(1-\alpha)^2}.$$

By importance sampling, it is not possible to modify the quantity $f_{\varphi(X)}(\xi^*)$ since it is an intrinsic constant which appears in the CLT (4.6) through the Jacobian matrix of h, where $h(\xi, C) := \mathbb{E}\left[H(\xi, C, X)\right]$ and $H(\xi, C, X) := (H_1(\xi, C, X), H_2(\xi, C, X))$.

Consequently, we are led to find the parameters θ^* and μ^* minimizing the two functionals:

$$Q_1(\theta,\xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p(X-\theta)}\right], \quad Q_2(\mu,\xi^*) := \mathbb{E}\left[(\varphi(X)-\xi^*)_+^2\,\frac{p(X)}{p(X-\mu)}\right]. \tag{4.23}$$

The I.S. procedure investigated in [6] consists in using the second I.S. algorithm based on (4.22), namely

$$\tilde{L}_3(\theta,\xi) := \mathbf{1}_{\varphi(\Psi(u)-\theta)\geq\xi}W(\theta,\Psi(u)),$$
$$\tilde{L}_4(\mu,\xi) := (\varphi(\Psi(u)-\mu)-\xi)_+^2\,e^{-(A+1)|\mu|^b}W(\mu,\Psi(u)).$$

where A is some positive real constant. The VaR-CVaR procedure is modified adaptively by the two I.S. algorithms

$$\xi_n = \xi_{n-1} - \gamma_n\tilde{L}_1(\xi_{n-1},\theta_{n-1},u_n) \tag{4.24}$$
$$C_n = C_{n-1} - \gamma_n\tilde{L}_2(\xi_{n-1},C_{n-1},\mu_{n-1},u_n). \tag{4.25}$$

where for all $\xi\in\mathbb{R},\ u\in[0,1]^q,\ \theta\in\mathbb{R}^d,\ \mu\in\mathbb{R}^d$,

$$\tilde{L}_1(\xi,\theta,u) := 1-\frac{1}{1-\alpha}\mathbf{1}_{\varphi(\Psi(u)+\theta)\geq\xi}\frac{p(\Psi(u)+\theta)}{p(\Psi(u))},$$
$$\tilde{L}_2(\xi,c,\mu,u) := c-\xi-\frac{1}{1-\alpha}\,(\varphi(\Psi(u)+\mu)-\xi)_+\,\frac{p(\Psi(u)+\mu)}{p(\Psi(u))}.$$

From a theoretical point of view, we didn't succeed in proving that the new procedure $(\xi_n,C_n,\theta_n,\mu_n)_{n\geq0}$ defined as above converges toward $(\xi_\alpha^*,C_\alpha^*,\theta_\alpha^*,\mu_\alpha^*)$. However, from a numerical point of view, we observe that this procedure converges and that the I.S. does accelerate the convergence of the initial VaR-CVaR algoritm defined by (4.9) and (4.10).

4.4 Numerical results

Tukey-Lambda distribution

The first example is a trivial case where we want to estimate the VaR and the CVaR of the Tukey-lambda distribution. This distribution is often used to identify an appropriate distribution. Although its probability density function does not have a simple form, the inverse of its distribution function, *i.e.* its quantile function is explicitly known and is given by

$$F^{-1}(p) = \frac{1}{\lambda}\left(p^\lambda-(1-p)^\lambda\right), \text{ for } \lambda\neq0, \quad \text{and} \quad F^{-1}(p) = \log(p)-\log(1-p), \text{ for } \lambda=0.$$

We devise the algorithm defined by (4.9) and (4.10) for three different values of $\alpha = 95\%,\ 99\%,\ 99.5\%$. Note that $\varphi\equiv Id$ and for $\lambda>1$, $\Psi\equiv F^{-1}$ is clearly lipschitz so that owing to Theorem 4.3.3, the VaR-CVaR quasi-stochastic approximation algorithm converges. Of course, this is just a toy-example since the explicit computation of ξ_α^* is trivial since $\xi_\alpha^* := F^{-1}(\alpha) = \frac{1}{\lambda}\left(\alpha^\lambda-(1-\alpha)^\lambda\right)$. However, note that C_α^* is not explicitly known. Concerning the uniformly distributed sequence

with low discrepancy, we use a Van Der Corput sequence, for more details about this sequence, we refer to [70].

The following simulation and results were achieved with $\lambda = 5$. The step sequence is defined for $n \geq 0$ by $\gamma_n = \frac{10}{n+200}$.

The results displayed in Table 4.1 correspond to the VaR_α, CVaR_α with both VaR and CVaR procedures using a low discrepancy sequence and a pseudo-random sequence. The columns VaR_{qmc} and CVaR_{qmc} correspond to the VaR and the CVaR estimates using (4.9) and (4.10). The columns VaR and CVaR correspond to the VaR and the CVaR estimate using the procedure developed in [6]. Although we don't have any *a priori* error bound, Figure 4.1 clearly indicate that it reduces the number of steps needed to obtain good approximates.

Number of steps	α	VaR_{qmc}	CVaR_{qmc}	VaR	CVaR
500	95%	0.1583	0.1804	0.1818	0.1992
	99%	0.1909	0.2088	0.2303	0.2131
	99.5%	0.1947	0.1725	0.1950	0.2169
1000	95%	0.1578	0.1809	0.1586	0.1862
	99%	0.1928	0.2072	0.1910	0.1973
	99.5%	0.1995	0.1841	0.2328	0.2396
10 000	95%	0.1562	0.1783	0.1509	0.1729
	99%	0.1940	0.2009	0.1913	0.1980
	99.5%	0.1908	0.2010	0.1909	0.2011

Table 4.1: The values of the VaR_α of the Tukey-lambda distribution are: $\xi^*_{0.95} = 0.15475$, $\xi^*_{0.99} = 0.19019$, $\xi^*_{0.995} = 0.19505$.

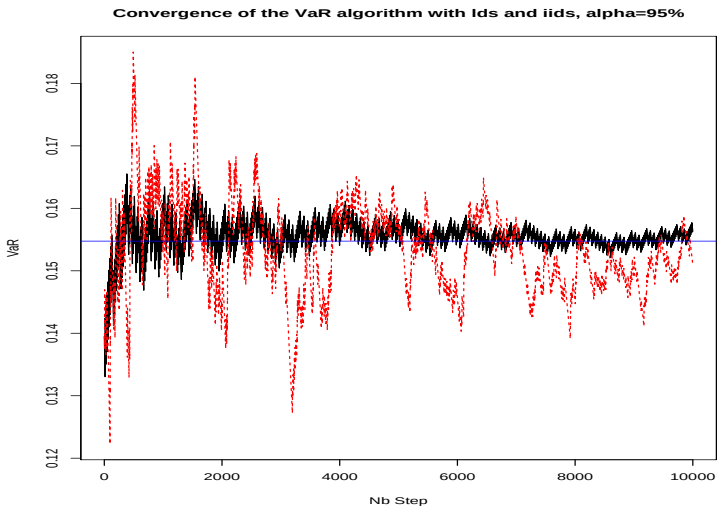


Figure 4.1: Convergence of the VaR procedure using a low-discrepancy sequence (normal lines) and a pseudo-random sequence (dashed lines) at level $\alpha = 95\%$. The horizontal line is the value of $\xi^*_{0.95} = 0.15475$.

Calls and Puts portfolio

We consider a portfolio composed of short positions in 10 calls and 10 puts on each of 5 underlying assets, all options having the same maturity 0.25 years with several strikes. The underlying assets are modeled as geometric Brownian Motions, all having a volatility of 20% and are assumed to be uncorrelated. The dimension d of the structural vector X is equal to 5. The procedure (4.9) and (4.10) is implemented with a sequence of some quasi-random normal numbers, namely

$$(X_n^k, X_n^{k+1}) = \left(\sqrt{-2\log(u_n^k)}\sin(2\pi u_n^{k+1}), \sqrt{-2\log(u_n^k)}\cos(2\pi u_n^{k+1})\right), \quad k = 1,\ 3,\ 5.$$

where $u_n = (u_n^k)_{1\le k\le 6}$, $n \ge 1$ is simply a regular 6-dimensional Sobol sequence (see e.g. [70]). We compare the performance of the algorithm (4.9) using a Sobol sequence (VaR(qmc)) and stochastic approximation for the computation of the VaR_α with (VaR(IS)) and without (VaR) variance reduction using pseudo-random numbers. The numerical results are reported in Table 4.2.

Number of steps	α	VaR	VaR(IS)	VaR(qmc)
1000	95%	305.4	317.1	330.4
	99%	493.9	490.1	411.3
	99.5%	291.5	540.1	507.3
	99.9%	471.8	484	534.1
10 000	95%	348.6	356.8	362.2
	99%	501.7	494.2	502.7
	99.5%	475.5	551.9	543.3
	99.9%	597.1	598.6	612.1
100 000	95%	364.1	361.1	362.5
	99%	507.3	502.9	506.9
	99.5%	553.3	554.1	555.1
	99.9%	648.2	652.5	657.4

Table 4.2: Portfolio 2 (QMC).

Spark spread Option We consider a portfolio composed of a spark spread option with a maturity of 1 year and where gas spot prices S_T^g and electricity spot prices S_T^e are modeled by two geometric brownian motions, all having a volatility of 40%. Electricity and gas initial spot prices are $S_0^e = 40$ \$/MWh and $S_0^g = 3$ \$/MMBTU (BTU: British Thermal Unit) with a Heat Rate equals $h_R = 10$ BTU/kWh and generation costs $C = 5$ \$/MWh. The two spot prices are uncorrelated. The loss can be written

$$L = \left(S_T^e - h_R S_T^g - C\right)_+ .$$

The dimension d of the structural vector X is equal to 2. In order to estimate $(\xi_\alpha^*, C_\alpha^*)$, we use the VaR-CVaR algorithm with adaptive I.S. as developed in Section 4.3.2.

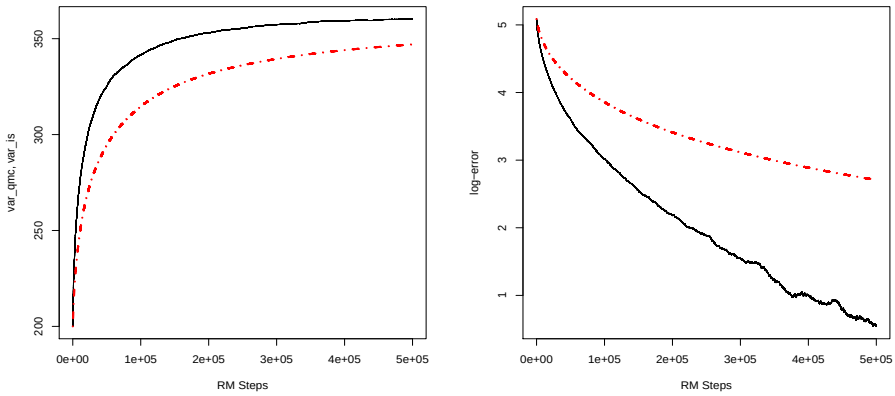


Figure 4.2: Fig 4.3.a (left): VaR using a Sobol sequence (normal line) and VaR using IS procedure with pseudo-random numbers (dotted line) for $\alpha = 95\%$. Fig 4.3.b (right): log-distance to the VaR_α estimate using a Sobol sequence (normal line) and using IS procedure (dotted line). *The VaR (QMC) converges faster than the VaR with IS. A good approximate of* VaR_α *is approximately 362 (*$\log(362) \approx 5.9$*).*

Number of steps	α	VaR(IS)	VaR(qmc)	CVaR(IS)	CVaR(qmc)
1000	95%	38.1	35.3	49.5	53.5
	99%	54.9	60	156.1	78.3
	99.5%	58.3	68.6	68.7	96.1
	99.9%	68.1	84.3	149.1	107.2
10 000	95%	41.8	37.6	53.5	54.5
	99%	55.3	62.4	89.1	79.1
	99.5%	66.4	73.7	83.1	88.1
	99.9%	81.7	85.2	116.4	108.3
100 000	95%	40.8	38.8	54.7	54.1
	99%	59.3	62.7	77.4	76.7
	99.5%	69.3	74.1	88.8	86.5
	99.9%	88.9	90.4	114.3	110.1

Table 4.3: Portfolio 3 (QMC).

Part II

CVaR Hedging using stochastic approximation

Chapter 5

CVaR hedging using quantization based stochastic approximation algorithm

Joint work with O. Bardou and G. Pagès.

Abstract: In this paper, we investigate a method based on risk minimization to hedge observable but non-tradable source of risk on financial or energy markets. The optimal portfolio strategy is obtained by minimizing dynamically the Conditional Value-at-Risk (CVaR) using three main tools: stochastic approximation algorithm, optimal quantization and variance reduction techniques (importance sampling (IS) and linear control variable (LCV)) as the quantities of interest are naturally related to rare events. As a first step, we investigate the problem of CVaR regression, which corresponds to a static portfolio strategy where the number of units of each tradable assets is fixed at time 0 and remains unchanged till time T. We devise a stochastic approximation algorithm and study its a.s. convergence and rate of convergence. Then, we extend to the dynamic case under the assumption that the process modelling the non-tradable source of risk and financial assets prices are Markov. Finally, we illustrate our approach by considering several portfolios in the incomplete energy market.

Keywords: VaR, CVaR, Stochastic Approximation, Robbins-Monro algorithm, Quantification.

5.1 Introduction

It is well known that in a complete financial market, an investor faced with a contingent claim can hedge perfectly on a finite horizon time T without any risk. However, from a practical standpoint, an agent would like to have a more realistic view of financial or energy markets which are intrinsically incomplete for many reasons (stochastic volatility, jumps, temperature dependance of prices on energy markets, ...). There is no exact replication to provide a unique price. Thus, pricing and hedging contingent claims in such a framework require new approaches. One may still price and hedge using a super-hedging criterion as studied in [26]. However, the price is often too high, actually, the trader can only hedge partially and often has to bear some risk of loss. Many authors studied pricing theory under a martingale measure which corresponds to an optimized criterion. For instance, one can refer to [32] for the minimal martingale measure, to [4], [34] and [67] for the minimal entropy martingale measure among others.

Another method widely used to adress this problem is based on expected utility maximization. Indeed, there is a huge litterature on hedging and pricing in incomplete markets using expected utility maximization method and utility indifference pricing. It consists in pricing an unhedgable claim so that the investor's utility remains unchanged between holding and not holding the contingent claim. We refer to [44], [47] and [27] among many others for some developments. Although, this approach has been studied for long, the main drawback for a pratician remains the lack of knowledge of his own utility function for hedging and pricing derivatives. Moreover, different agents may price and hedge a contingent claim differently according to their own risk preference so that it has little acceptance in practice.

In this article, we propose an alternative method based on risk minimization using stochastic approximation algorithm. To be more precise, we focus on minimizing dynamically the Conditional Value-at-Risk (CVaR). The CVaR is strongly linked to the famous risk measure called Value-at-Risk (VaR) which is certainly the most widely used risk measure in the practice of risk management. By definition, the VaR at level $\alpha \in (0,1)$ (VaR_α) of a given portfolio loss distribution is the lowest amount not exceeded by the loss with probability α (usually $\alpha \in [0.95, 1)$). The Conditional Value-at-Risk at level α (CVaR_α) is the conditional expectation of the portfolio losses beyond the VaR_α level. Compared to VaR, the CVaR is known to have better properties. Risk measures of this type were introduced in [3] and have been shown to share basic coherence properties (which is not the case of VaR_α). The extension to convex risk measures were introduced and extensively studied in [31].

Pricing and hedging using risk measures is a recent approach which has been investigated by many authors. Barrieu and El Karoui in [9] developped a risk minimization problem to hedge non-tradable risk on financial market using convex risk measures. Hedging strategy which maximizes the probability of successfull hedge is studied in [30] as an alternative to super-hedging strategy which requires a large amount of initial capital.

In [77], a portfolio optimization method which calculates the VaR_α and optimizes the CVaR_α using a linear programming approach is developed. Portfolio strategies with a low CVaR_α necessarily have a low VaR_α. The method first consists in generating loss scenarios and then in introducing them as constraints in the linear programming problem. The main drawback is that the dimension (number of

constraints) of the linear programming problem to be solved is equal to the number of simulated scenarios so that this approach turns out to have strong limitations in practice. In our approach, we are not limited by the number of simulated scenarios.

We consider an energy (or financial) market operating at discrete trading dates $t_0 = 0 < t_1 < \cdots < t_M = T$. We have d assets available for trade with price process $X = \left(X^1, \cdots, X^d\right)$ and $X^i = \left(X^i_{t_\ell}\right)_{0\leq \ell \leq M}$ for $i = 1, \cdots, d$. We will denote X_ℓ for X_{t_ℓ}. For simplicity, we assume that the risk free rate is equal to zero. The portfolio loss (or the payoff of a financial instrument) with maturity T is described by an $\mathbb{R}$-valued random variable L defined on a probability space $(\Omega, \mathcal{G}, \mathbb{P})$. In our framework, the source of market incompleteness comes from the presence in L of a state process Z that is *observable* but *not available* for trade. Thus, it induces a source of risk that is not completely hedgable. Typically, in the electricity market, the loss L suffered by an energy company may be due to an anormal annual electricity (or gas) anormal consumption. This consumption depends on the temperature, which is an *observable* but non tradable source of risk. In this example the process $(Z_\ell)_{1\leq \ell \leq M}$ can be considered as the temperature which may influence not only the loss but the assets available for trade, *i.e.* electricity prices of spot and forward contracts (which are in this example the only available assets for hedge). More generally, this kind of dependance with respect to an observable but non available source of risk is a particularly relevant source of incompleteness in financial and energy markets (stochastic volatility, default time, temperature for energy derivatives, weather contracts, ...). The probability space is equiped with a filtration $\mathbb{G} = (\mathcal{G}_\ell)_{0\leq \ell \leq M}$. Intuitively, $\mathcal{G}_\ell$ represents the observable information at time t_ℓ by all investors, so that $\mathcal{G}_\ell = \sigma\left\{X_i, Z_i; 0 \leq i \leq \ell\right\}$.

In order to reduce its risk (or hedge the contingent claim), the holder of the portfolio uses a dynamic self-financed strategy represented by a d-dimensional predictable process $\theta = (\theta_\ell)_{0\leq \ell < M}$, where $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{G}_\ell, \mathbb{P})$ ($L^0_{\mathbb{R}^d}(\mathcal{G}_\ell, \mathbb{P})$ denotes the space of all $\mathcal{G}$-mesurable and $\mathbb{P} - a.s.$ finite random variables with values in $\mathbb{R}^d$). In such a strategy, we may regard θ_ℓ as the number of shares invested in the stock at time t_ℓ. The gains from a self-financed trading strategy θ with an initial investment of 0 are described by the discrete stochastic integral $\sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell$, where we denote by ΔX_ℓ the increments $X_\ell - X_{\ell-1}$. Throughout the paper, we will use the following main assumptions

Assumption 5.1.1. *The process $(X_\ell)_{0\leq \ell \leq M}$ is a $(\mathbb{G}, \mathbb{P})$-martingale,*

and

Assumption 5.1.2. *The process $(X_\ell, Z_\ell)_{0\leq \ell \leq M}$ taking its values in $\mathbb{R}^d \times \mathbb{R}^q$ is Markovian with respect to the filtration $\mathbb{G}$.*

In energy markets, there are two kind of assets available for trade: *day-ahead* or spot and *forward* contracts. However, due to the impossibility to store electricity or high costs to store gas, it is not possible to use the electricity and the gas spot contracts in a self-financed trading strategy. Consequently, one often has to rely on forward contracts which are the only assets available for trade. Forward price dynamic is often modelized using Heath-Jarrow-Morton approach directly under the historical probability $\mathbb{P}$ so that assumption 5.1.1 may be satisfied in practice.

The basic problem for the holder of the portfolio is to find the optimal self-financed strategy θ_α^* which minimizes the residual risk of the portfolio's losses, *i.e.* the solution of the following minimization problem

$$\inf_{\theta\in\mathcal{A}_\mathcal{G}} \mathrm{CVaR}_\alpha\left(L-\sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell\right)^{1}, \tag{5.1}$$

where $\mathcal{A}_\mathcal{G} = \left\{\theta=(\theta_\ell)_{0\leq\ell\leq M-1} \mid \theta_\ell\in L^0_{\mathbb{R}^d}(\mathcal{G}_\ell,\mathbb{P}),\ \ell=0,\cdots,\ M-1\right\}$ is the set of admissible strategies, that is minimizing the residual risk of the portfolio risk profile over all self-financed strategies.

A natural question which arises is how to measure dynamically the risk of the considered portfolio in this context. To measure the risk at a given time t_ℓ, we introduce in a quite natural way and for the first time to our knowledge, the definition of a dynamic version of the CVaR that will be denoted $\mathcal{G}_\ell$-CVaR based on the Rockafellar & Uryasev's static representation of the CVaR. In order to estimate at time 0, this random risk measure, which reads as a conditional expectation, we use integration cubature formula based on optimal quantization.

For many reasons (transaction costs, difficulties to store energy assets, ...), the holder of the portfolio may not want to trade every day but may be only interested by a rough hedge to reduce its risk. Consequently, we firstly investigate one step self-financed strategies. Decided at time t_{ℓ_0} such strategy is obtained by setting $\theta_k\equiv\theta_{\ell_0}$, for $k=\ell_0,\cdots,M-1$. Consequently, a one-step portfolio strategy decided at time t_{ℓ_0} is an $\mathbb{R}^d$-valued random variable $\theta_{\ell_0}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0},\mathbb{P})$. The investor risk at time t_{ℓ_0} can be measured by the quantity $\mathcal{G}_{\ell_0}$-$\mathrm{CVaR}_\alpha\left(L-\theta_{\ell_0}.(X_M-X_{\ell_0})\right)^2$ which is only known at time t_{ℓ_0}. However, the investor can estimate this quantity at time 0 by numerically computing $\mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-}\mathrm{CVaR}_\alpha\left(L-\theta_{\ell_0}.(X_M-X_{\ell_0})\right)\right]$. This quantity is a forward risk, *i.e.* it is the best estimation at time 0 of the risk at time t_{ℓ_0} while the quantity $\mathrm{CVaR}_\alpha\left(L-\theta_{\ell_0}.(X_M-X_{\ell_0})\right)$ represents the risk at time 0. Consequently, there are two optimization problems.

The first one is to minimize the forward risk, *i.e.* the expectation of the risk profile measured at time t_{ℓ_0} of the portfolio losses using a self-financed one step portfolio strategy starting from an initial wealth of 0

$$\inf_{\theta_{\ell_0}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0},\mathbb{P})} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-}\mathrm{CVaR}_\alpha\left(L-\theta_{\ell_0}.(X_M-X_{\ell_0})\right)\right]. \tag{5.2}$$

The second one consists in minimizing the risk measured at time 0 (*i.e.* we use a static CVaR criterion) of the portfolio losses using a self-financed one step portfolio strategy starting from an initial wealth of 0

$$\inf_{\theta_{\ell_0}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0},\mathbb{P})} \mathrm{CVaR}_\alpha\left(L-\theta_{\ell_0}.(X_M-X_{\ell_0})\right). \tag{5.3}$$

The VaR_α and the CVaR_α are disymetric risk measures unlike standard deviation. By CVaR hedging we aim at modifying the shape of the loss distribution

[1] We consider the general definition of expectation of a random variable Y, *i.e.* the quantity $\mathbb{E}[Y]$ exists as soon as $\mathbb{E}[Y_+]<+\infty$ or $\mathbb{E}[Y_-]<+\infty$.

[2] We consider the general definition of conditional expectation of a random variable Y, *i.e.* the quantity $\mathbb{E}[Y|\mathcal{G}_{\ell_0}]$ as soon as $\mathbb{E}[Y_+|\mathcal{G}_{\ell_0}]<+\infty$ or $\mathbb{E}[Y_-|\mathcal{G}_{\ell_0}]<+\infty$

L, *i.e.* we reduce the right-hand side of the distribution which corresponds to high loss greater than the left-hand side which corresponds to small losses or potential gains. That is the main difference between CVaR hedging and hedging by means of a quadratic criterion as developed in [33] and [83] among others.

Under a Markovian framework, *i.e.* under Assumptions 5.1.1 and 5.1.2, we propose a stochastic approximation algorithm to compute the optimal self-financed portfolio strategy θ_α^* solution of (5.3), (5.2) and (5.1) (and both the VaR and the CVaR of the resulting portfolio).

However, in the case of dynamic self-financed strategies framework, when the number of trading dates M is too large (say $M \geq 10$, in practice) or when the dimension of the process (X, Z) is too large, the proposed algorithm to solve (5.1) turns out to be numerically untractable. We develop other approaches based on some majorations of the objective function of (5.1) in order to approximate the optimal solution.

All proposed algorithms are built on some Rockafellar & Uryasev's representation of the CVaR and spatial discretization of the process $(X_\ell, Z_\ell)_{0\leq\ell\leq M}$ using optimal vector quantization. This leads us to devise a global Robbins-Monro (RM) procedure to estimate all the quantities of interest. This kind of idea has already been used in [6] to propose an algorithm which simultaneously computes both the VaR and the CVaR. The estimator provided by the algorithm satisfies the standard Central Limit Theorem (CLT) for recursive stochastic algorithm. However, the proposed algorithm is just a first building block. When α is close to 1 (otherwise the original procedure behaves well), VaR and CVaR are fundamentaly related to rare events. As a matter of fact, in this kind of probem, we are interested in hedging extreme events, *i.e.* events that are observed with a very small probability (usually less than 5%, 1% or even 0.1%) thus we obtain few significant scenarios to update our estimates. As a crucial improvement, we need to introduce a recursive variance reduction method. To compute more accurate estimates, it is necessary to generate more samples in the area of interest, the tail of the distribution. A natural tool used in this situation is importance sampling (IS). Following the IS procedure developed in [62], which has already been used in [6] for the estimation of the VaR and the CVaR, our IS parameters are optimized adaptively by a companion (unconstrained) RM algorithm which is combined with our first procedure. We also propose another variance reduction method based on a linear control variable which can be used alone when IS is not necessary or can be combined with the IS algorithm. It dramatically accelerates the convergence of the original procedure. The weak convergence rate of the resulting procedure is ruled by a CLT with optimal rate and minimal variance.

The paper is organized as follows: in Section 5.2, we present the dynamic version of the CVaR and develop some fundamental theoretical results on the $\mathcal{G}$-CVaR and CVaR hedging. This will allow us to devise a RM algorithm. Section 5.3 is devoted to numerical aspects of CVaR hedging. We show how to devise a RM algorithm to compute the optimal strategy with its associated VaR and CVaR. We establish its *a.s.* convergence and rate of convergence. In order to approximate conditional expectation, we rely on optimal vector quantization. We present our several algorithms to approximate the optimal strategy solution of (5.1). In Section 5.4, we introduce and develop the two variance reduction tools in the strategy framework. We show how it modifies the asymptotic variance of the CLT. Finally, Section 5.5 is devoted to numerical examples. We focus on the energy market which is known to

be incomplete. We propose several portfolios to challenge the algorithm and display dynamic CVaR estimations.

Acknowledgments: I would like to thank Sandrine Bouthemy and Sophie Auchapt for stimulating discussions.

Notations: • $|.|$ will denote the canonical Euclidean norm on $\mathbb{R}^d$, $u.v$ will denote the canonical inner product of the two column vector $u,\ v \in \mathbb{R}^d$ and u^T denotes the transpose of the column vector $u \in \mathbb{R}^d$.
• $\xrightarrow{\mathcal{L}}$ will denote the convergence in distribution and $\xrightarrow{a.s.}$ will denote the almost sure convergence.
• $x_+ := \max(0, x)$ will denote the positive part function.
• $L^p(\mathbb{P})$ will denote the sub-space of random variable U such that $\left(\mathbb{E}\left[|U|^p\right]\right)^{1/p} < +\infty$.
• $L^p(du)$ will denote the sub-space of function f such that $\left(\int |f|^p du\right)^{1/p} < +\infty$.

5.2 Theoretical aspects of CVaR hedging

5.2.1 Definitions and preliminaries

We start this section by briefly recalling the definitions of the VaR and the CVaR (for more details, we refer to [6]). Then, we introduce the notion of dynamic CVaR that will be fundamental throughout the paper. To measure the risk associated to a loss (or a short position on the contingent claim with payoff) L, one usually considers the VaR at level $\alpha \in (0,1)$ *i.e.* the lowest α-quantile of the distribution L

$$\mathrm{VaR}_\alpha(L) := \inf\left\{\xi \in \mathbb{R} \mid \mathbb{P}\left(L \leq \xi\right) \geq \alpha\right\}.$$

We assume that the distribution function of L is continuous (*i.e.* with no atom) so that the VaR is the lowest solution of the equation:

$$\mathbb{P}\left(L \leq \xi\right) = \alpha.$$

If the distribution function is (strictly) increasing, the above equation has a unique solution, otherwise there may (infinitely) more. In fact, in what follows, we will consider that *any* solution of the previous equation is the $\mathrm{VaR}_\alpha(L)$. Another risk measure commonly used to provide information about the tail of the distribution of L is the Conditional Value-at-Risk (at level α). Assuming that $L \in L^1(\mathbb{P})$, it is defined by:

$$\mathrm{CVaR}_\alpha(L) := \mathbb{E}\left[L | L \geq \mathrm{VaR}_\alpha(L)\right].$$

The next proposition shows that these two quantities are solutions to a convex optimization problem which value function can be represented as an expectation, as pointed out in [77]. It has already been used in [6] to devise a RM algorithm to compute both the VaR and the CVaR. We briefly recall this important result in order to justify the definition of the dynamic CVaR.

Proposition 5.2.1. *Suppose that the distribution function of L is continuous and that $L \in L^1(\mathbb{P})$. Let V be the function defined on $\mathbb{R}$ by:*

$$V(\xi) = \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\xi)_+\right]. \tag{5.4}$$

Then, the function V is convex, Lipschitz continuous, differentiable and $\mathrm{VaR}_\alpha(L)$ *is any point of the set*

$$\arg\min V = \{\xi \in \mathbb{R} \mid V'(\xi) = 0\} = \{\xi \in \mathbb{R} \mid \mathbb{P}(L \leq \xi) = \alpha\},$$

where V' denotes the derivative of V. This derivative V' can in turn be represented as an expectation by

$$\forall\, \xi \in \mathbb{R}, \quad V'(\xi) = \mathbb{E}\left[1 - \frac{1}{1-\alpha}\mathbf{1}_{\{L \geq \xi\}}\right]. \tag{5.5}$$

Furthermore,

$$\mathrm{CVaR}_\alpha(L) = \min_{\xi \in \mathbb{R}} V(\xi). \tag{5.6}$$

We refer to [77] or [6] for a proof. Now we are in position to define the dynamic CVaR. We consider a sub σ-field $\mathcal{F} \subseteq \mathcal{G}$, representative of the information observable by all investors. Given the above result concerning the CVaR, it is quite natural to define the $\mathcal{G}$-CVaR according to the following definition.

Definition 5.2.1. Let $L \in L^1(\mathbb{P})$. The $\mathcal{F}$-CVaR is a random risk measure defined by

$$\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L) := \operatorname*{ess\,inf}_{\xi \in L^0(\mathcal{F})} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\xi)_+ \,|\mathcal{F}\right].$$

By construction, it is straightforward that it satisfies the following *coherence* properties

1. Sub-additivity: for every $L,\ L' \in L^1(\mathbb{P})$, $\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L+L') \leq \mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L) + \mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L')$.

2. Positive homogeneity: If $\lambda \in L^0(\mathcal{F})$ with $\lambda \geq 0$ *a.s.*, $\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(\lambda L) = \lambda \times \mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L)$.

3. Translation invariance: for all $Z \in L^0(\mathcal{F})$, $\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L+Z) = Z + \mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L)$.

4. Monotonicity: for every $L,\ L' \in L^1(\mathbb{P})$ such that $L \leq L'$, $\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L) \leq \mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L')$.

When $\mathcal{F} = \{\varnothing, \Omega\}$, the $\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L)$ coincides with the usual $\mathrm{CVaR}_\alpha(L)$. To estimate $\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L)$ at time 0, one may compute the quantity $\mathbb{E}\left[\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L)\right]$ which is still a coherent risk measure in the sense of [3].

5.2.2 General properties

In this section, we state some useful properties satisfied by the $\mathcal{G}_\ell\text{-}\mathrm{CVaR}_\alpha$.

If one aims at measuring the risk at time t_ℓ of his financial strategy $\theta \in \mathcal{A}$ started at time $t_0 = 0$ using a CVaR criterion, one has to compute $\mathcal{G}_\ell\text{-}\mathrm{CVaR}_\alpha\left(L - \sum_{p=1}^{M} \theta_{p-1}.\Delta X_p\right)$, which is only known at time t_ℓ. It is natural for the holder of the portfolio to ask how the risk evolves with time until maturity. Next result shows that the $\mathcal{G}_\ell$-CVaR risk measure is time consistent, *i.e.* the risk of any position decreases with time.

Proposition 5.2.2. *We set $M = +\infty$ for this result. Let $Y \in L^1(\mathcal{G}_\infty, \mathbb{P})$ where $\mathcal{G}_\infty = \vee_\ell \mathcal{G}_\ell$.*
The sequence $(\mathcal{G}_\ell\text{-CVaR}_\alpha(Y))_{1\le\ell\le M}$ is a $\mathbb{G}$-supermartingale. Moreover, it satisfies,

$$\mathcal{G}_n\text{-CVaR}_\alpha(Y) \xrightarrow{a.s.} Y, \quad \text{as } n \to +\infty.$$

Proof. First note that for $\ell = 1, \cdots, M$,

$$\mathcal{G}_\ell\text{-CVaR}_\alpha(Y) = \operatorname*{ess\,inf}_{\xi\in L^0(\mathcal{G}_\ell)} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y-\xi)_+ |\mathcal{G}_\ell\right] \le \frac{1}{1-\alpha}\mathbb{E}\left[Y_+|\mathcal{G}_\ell\right] \in L^1(\mathbb{P}),$$

and by Jensen's inequality,

$$\mathbb{E}\left[Y|\mathcal{G}_\ell\right] = \operatorname*{ess\,inf}_{\xi\in L^0(\mathcal{G}_\ell)} \xi + \frac{1}{1-\alpha}\left(\mathbb{E}\left[Y|\mathcal{G}_\ell\right]-\xi\right)_+ \le \mathcal{G}_\ell\text{-CVaR}_\alpha(Y), \tag{5.7}$$

so that, $\mathcal{G}_\ell$-CVaR$_\alpha(Y) \in L^1(\mathbb{P})$. Then, by definition, we have

$$\mathcal{G}_\ell\text{-CVaR}_\alpha(Y) \le \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y-\xi)_+ |\mathcal{G}_\ell\right], \quad \text{for all } \xi \in L^0(\mathcal{G}_{t_{\ell-1}}),$$

which implies that

$$\mathbb{E}\left[\mathcal{G}_\ell\text{-CVaR}_\alpha(Y)|\mathcal{G}_{\ell-1}\right] \le \operatorname*{ess\,inf}_{\xi\in L^0(\mathcal{G}_{\ell-1})} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y-\xi)_+ |\mathcal{G}_{\ell-1}\right] = \mathcal{G}_{\ell-1}\text{-CVaR}_\alpha(Y).$$

Consequently, the sequence $(\mathcal{G}_\ell\text{-CVaR}_\alpha(Y))_{1\le\ell\le M}$ is a $\mathbb{G}$-supermartingale.
Now, owing to (5.7), for $n \ge 1$

$$\left(\mathcal{G}_n\text{-CVaR}_\alpha(Y)\right)_- \le \left(\mathbb{E}\left[Y|\mathcal{G}_n\right]\right)_- \le \left(\mathbb{E}\left[Y_-|\mathcal{G}_n\right]\right) \le \mathbb{E}\left[|Y|\,|\mathcal{G}_n\right],$$

and

$$\sup_{n\ge0}\mathbb{E}\left[\left(\mathcal{G}_n\text{-CVaR}_\alpha(Y)\right)_-\right] \le \mathbb{E}\left[|Y|\right] < +\infty.$$

Doob's martingale convergence theorem implies that the sequence $\left(\mathcal{G}_n\text{-CVaR}_\alpha(Y)\right)_{n\ge1}$ *a.s.* converges toward $\tilde{Y}_\infty \in L^1(\mathbb{P})$. Now, from the first inequality and the *a.s.* convergence of the sequence $\left(\mathbb{E}\left[Y|\mathcal{G}_n\right]\right)_{n\ge1}$ toward $\mathbb{E}\left[Y|\mathcal{G}_\infty\right] = Y$ (the convergence also holds in L^1), we get

$$\tilde{Y}_\infty \ge Y.$$

On the other hand, for every $n \ge 1$

$$\mathcal{G}_n\text{-CVaR}_\alpha(Y) \le \mathbb{E}\left[Y|\mathcal{G}_n\right] + \frac{1}{1-\alpha}\mathbb{E}\left[\left(Y-\mathbb{E}\left[Y|\mathcal{G}_n\right]\right)_+ |\mathcal{G}_n\right],$$

so that, for every $n \ge m \ge 1$ and every $A \in \mathcal{G}_m$,

$$\mathbb{E}\left[\mathbf{1}_A\mathcal{G}_n\text{-CVaR}_\alpha(Y)\right] \le \mathbb{E}\left[\mathbf{1}_A\left(Y + \frac{1}{1-\alpha}\left(Y-\mathbb{E}\left[Y|\mathcal{G}_n\right]\right)_+\right)\right]. \tag{5.8}$$

It follows from Fatou's Lemma that

$$\mathbb{E}\left[\mathbf{1}_A\tilde{Y}_\infty\right] = \mathbb{E}\left[\mathbf{1}_A\underline{\lim}_n\mathcal{G}_n\text{-CVaR}_\alpha(Y)\right] \le \underline{\lim}_n\mathbb{E}\left[\mathbf{1}_A\mathcal{G}_n\text{-CVaR}_\alpha(Y)\right],$$

since $\mathcal{G}_n\text{-CVaR}_\alpha(Y) \geq \mathbb{E}[Y|\mathcal{G}_n]$, *a.s.*, for every $n \geq 1$ and $\mathbb{E}[Y|\mathcal{G}_n]$ converges in $L^1(\mathbb{P})$. Now $(Y - \mathbb{E}[Y|\mathcal{G}_n])_+ \overset{L^1(\mathbb{P})}{\longrightarrow} 0$ which shows that

$$\overline{\lim}_n \mathbb{E}\left[\mathbf{1}_A\left(Y + \frac{1}{1-\alpha}(Y - \mathbb{E}[Y|\mathcal{G}_n])_+\right)\right] \leq \mathbb{E}[\mathbf{1}_A Y].$$

Combining these inequalities with (5.8) yields

$$\forall\, m \geq 1,\ \forall\, A \in \mathcal{G}_m,\ \ \mathbb{E}\left[\mathbf{1}_A \tilde{Y}_\infty\right] \leq \mathbb{E}[\mathbf{1}_A Y],$$

which in turn implies that

$$\tilde{Y}_\infty \leq Y.$$

This completes the proof. □

This result naturally implies that the sequence $(\mathbb{E}[\mathcal{G}_\ell\text{-CVaR}_\alpha(Y)])_{1\leq \ell \leq M}$ is non-increasing, thus the average risk (hopefully) decreases with time for any strategy $\theta \in \mathcal{A}$. The result concerning the convergence of the supermartingale is quite intuitive. If the loss of the considered portfolio satisfies $L \in \mathcal{G}_M$ (as it is the case in our modelization) then the average risk associated to this position decreases toward the average loss itself.

Another useful result concerns the supermartingale property of the hedged portfolio.

Corollary 5.2.3. *Suppose that $L \in L^1(\mathbb{P})$ and that there exists $p' > 1$ such that $\Delta X_\ell \in L^{p'}(\mathbb{P})$ for $\ell = 1, \cdots, M$. Let $\theta \in \mathcal{A}$ such that $\theta_\ell \in L^p(\mathbb{P})$ with $p = \frac{p'}{p'-1}$. Then,*

$$\left(\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right)_{0\leq k\leq M} \quad \textit{is a supermartingale}$$

and satisfies, for every $k \in \{0, \cdots, M-1\}$,

$$\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) = \mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) - \sum_{\ell=1}^{\ell} \theta_{\ell-1}\Delta X_\ell. \tag{5.9}$$

Proof. Hölder's inequality implies that $\sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell \in L^1(\mathbb{P})$ so that in view of the definition of the $\mathcal{G}_k$-CVaR, $\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) \in L^1(\mathbb{P})$. Now by the change of variable, $\xi = \tilde{\xi} + \sum_{\ell=1}^{k} \theta_{\ell-1}\Delta X_\ell$, we have

$$\begin{aligned}
\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) &= \operatorname*{ess\,inf}_{\xi \in L^0(\mathcal{G}_k)} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+ \middle| \mathcal{G}_k\right] \\
&= \sum_{\ell=1}^{k} \theta_{\ell-1}\Delta X_\ell \\
&\quad + \operatorname*{ess\,inf}_{\tilde{\xi} \in L^0(\mathcal{G}_k)} \tilde{\xi} + \frac{1}{1-\alpha}\mathbb{E}\left[\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell - \tilde{\xi}\right)_+ \middle| \mathcal{G}_k\right] \\
&= \sum_{\ell=1}^{k} \theta_{\ell-1}\Delta X_\ell + \mathcal{G}_\ell\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell\right).
\end{aligned}$$

□

In particular, if Assumption 5.1.1 is satisfied (5.9) implies that for every $k \in \{0, \cdots, M-1\}$

$$\mathbb{E}\left[\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right] = \mathbb{E}\left[\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right],$$

which means that the mean estimate at time 0 of the risk at time t_k does not depend on the decisions taken prior to time t_k. This property follows from the fact that the hedging strategy is self-financed.

5.2.3 CVaR hedging using a one step self financed strategy

In this section, we adress the two problems (5.2) and (5.3), that is hedging a contingent claim using a one step strategy starting with an initial investment of 0 and a CVaR or a $\mathcal{G}_{\ell_0}$-CVaR criterion at a fixed time t_{ℓ_0}.

By one step strategy decided at time t_{ℓ_0}, we mean that the investor is restricted to rebalance its portfolio only once at time $t_\ell \in \{t_0, \cdots, t_{M-1}\}$. By a one step static strategy, we mean that the investor uses a one step strategy decided at time $t_0 = 0$.

This case of study is interesting since in energy markets, practiciens may be interested only by a rough hedge of their loss using only few forward contracts, especially when dealing with physical assets like gas storage or power plant. Moreover, theoretical results in the dynamic framework will be built on similar ideas used in this section.

Without loss of generality, we can suppose that the market operates with only two dates t_{ℓ_0} and $T = t_M$. We will denote X for $X_M - X_{\ell_0}$. Actually, we use a general σ-algebra $\mathcal{F} \subseteq \mathcal{A}$ with the possibility of setting $\mathcal{F}$ to $\mathcal{G}_{\ell_0}$ with $\ell_0 = 0, \cdots, M-1$. Consequently, we consider the two more general problems

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F}, \mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right], \tag{5.10}$$

and

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F}, \mathbb{P})} \text{CVaR}_\alpha\left(L - \theta.X\right). \tag{5.11}$$

Note that (5.11) can be written

$$\inf_{\xi \in L^0_{\mathbb{R}}(\mathcal{F}, \mathbb{P})} \inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F}, \mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right] \tag{5.12}$$

so that, in a first step, one may adress the stochastic optimization problem

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F}, \mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right]. \tag{5.13}$$

Up to the change of variable $L := L - \xi$, we can suppose that $\xi = 0$ and $\alpha = 0$ so that, without loss of generality, the problem (5.13) is equivalent to minimizing the short fall risk

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F}, \mathbb{P})} \mathbb{E}\left[\left(L - \theta.X\right)_+\right]. \tag{5.14}$$

First, we will show that there exists an optimal one step trading strategy $\tilde{\theta}$ solution to (5.14) thus for all $\xi \in \mathbb{R}$ there exists $\theta^*_\alpha(\xi)$ solution to (5.13). Finally, we will come back to (5.12) and deduce the existence of an optimal ξ^*_α solution of (5.12)

Now in order to derive the existence of solutions to (5.10) and (5.11), we assume the existence of a regular conditional distribution of the couple (L, X) given $\mathcal{F}$ denoted by $\Pi(\mathrm{d}y, \mathrm{d}x) = \Pi(\omega, \mathrm{d}y, \mathrm{d}x)$ and we make the following assumptions on the conditionnal distribution of the couple (L, X).

Assumption 5.2.4 (Static Case)**.**

i) The distribution of L and X satisfies $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $X \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$.

ii) $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathbb{E}\left[(u.X)_+ \mid \mathcal{F}\right] > 0$ *a.s.*

Assumption 5.2.5 (Forward Case)**.**

i) The distribution of L and X satisfies $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $X \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$.

ii) $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathcal{F}\text{-CVaR}_\alpha\left(u.X\right) > 0$ *a.s.*

The following proposition is the key result to solve our optimization problem. The proof is postponed to an appendix and relies on classical arguments from stochastic control theory.

Proposition 5.2.6. *Let V_f and V_s be the two functions defined respectively on $\Omega \times \mathbb{R} \times \mathbb{R}^d$ and $\Omega \times \mathbb{R}^d$ by*

$$V_f(\omega, \xi, \theta) = \int v_f(\xi, \theta, y, x)\Pi(\omega, \mathrm{d}x, \mathrm{d}y), \tag{5.15}$$

$$V_s(\omega, \xi, \theta) = \int v_s(\theta, y, x)\Pi(\omega, \mathrm{d}x, \mathrm{d}y) \tag{5.16}$$

where

$$v_f(\xi, \theta, y, x) = \xi + \frac{1}{1-\alpha}\left(y - \theta.x - \xi\right)_+, \tag{5.17}$$

and

$$v_s(\theta, y, x) = \left(y - \theta.x\right)_+, \tag{5.18}$$

Then, we have

i) Static Risk: *Suppose that Assumption 5.2.4 is satisfied. Then, for all $\omega \in \Omega$, the function $V_s(\omega, .)$ is convex, lipschitz continuous and $\lim_{|\theta|\to+\infty} V_s(\omega, \theta) = +\infty$. Moreover, we have*

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[(L - \theta.X)_+\right] = \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[(L - \theta.X)_+ \middle| \mathcal{F}\right]\right], \tag{5.19}$$

and

$$\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[(L - \theta.X)_+ \middle| \mathcal{F}\right](\omega) = \min_{\theta \in \mathbb{R}^d} V_s(\omega, \theta). \tag{5.20}$$

ii) Forward Risk: *Suppose that Assumption 5.2.5 is satisfied. Then, for all* $\omega \in \Omega$, *the function* $V_f(\omega, ., .)$ *is convex, continuous and for all* $\xi \in \mathbb{R}$, $\lim_{|(\xi,\theta)|\to+\infty} V_f(\omega, \xi, \theta) = +\infty$. *Moreover, we have*

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right] = \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ \xi \in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ \middle| \mathcal{F}\right]\right] \tag{5.21}$$

and

$$\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ \xi \in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ \middle| \mathcal{F}\right](\omega) = \min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\omega, \xi, \theta). \tag{5.22}$$

Remark 4. *The non-degeneracy Assumptions 5.2.4 ii) and 5.2.5 ii) can be replaced by the stronger assumption:*

- $\mathbb{E}\left[X \mid \mathcal{F}\right] = 0$ *and* $\mathbb{E}\left[XX^T \mid \mathcal{F}\right]$ *is a.s. positive definite in* $\mathcal{S}(d, \mathbb{R})$.
- *The conditional distribution of* X *given* $\mathcal{F}$ *is continuous (no affine hyperplane has positive mass).*

Indeed, for $\omega \in \Omega$, *we have* $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),|u|=1} \mathcal{F}\text{-CVaR}_\alpha\left(u.X\right)(\omega) = \inf_{\xi\in\mathbb{R}, u\in\mathcal{S}_d(0,1)} V_f(\omega, \xi, u)$ *where* $\mathcal{S}_d\left(0,1\right) := \left\{u \in \mathbb{R}^d \mid |u| = 1\right\}$ *denotes the (compact) unit sphere. Furthermore, since the function* $v_f(\xi, ., y, x)$ *is Lipschitz continuous for all* $\xi, y \in \mathbb{R}$, $x \in \mathbb{R}^d$, *it follows that for any* $u, u' \in \mathcal{S}_d\left(0,1\right)$,

$$\begin{aligned}\left|\inf_{\xi\in\mathbb{R}} \int v_f\left(\xi, u, 0, x\right)\Pi(\mathrm{d}x, \mathrm{d}y) - \inf_{\xi\in\mathbb{R}} \int v_f\left(\xi, u', 0, x\right)\Pi(\mathrm{d}x, \mathrm{d}y)\right| & \\ \leq \sup_{\xi\in\mathbb{R}} \left|\int \left(v_f(\xi, u, y, x) - v_f(\xi, u', y, x)\right)\Pi(\mathrm{d}x, \mathrm{d}y)\right| & \\ \leq \frac{|u-u'|}{1-\alpha} \int |x|\Pi(\mathrm{d}x, \mathrm{d}y),\ a.s. &\end{aligned}$$

Consequently, for all $\omega \in \Omega$, *the function* $u \mapsto \inf_{\xi\in\mathbb{R}} V_f(\omega, \xi, u)$ *is continuous on* $\mathcal{S}_d\left(0,1\right)$. *Thus, it remains to check that for all* $u \in \mathcal{S}_d\left(0,1\right)$, $\inf_{\xi\in\mathbb{R}} V_f(\omega, \xi, u) > 0$, *knowing that* $\mathbb{E}\left[X \mid \mathcal{F}\right] = 0$ *a.s. Proposition 5.2.1 implies that there exists* ξ^*_α *such that* $\inf_{\xi\in\mathbb{R}} V_f(\omega, \xi, u) = V_f(\omega, \xi^*_\alpha, u)$. *There are three cases to check:*

- *if* $\xi^*_\alpha > 0$, *then it is straighforward that* $V_f(\omega, \xi^*_\alpha, u) \geq \xi^*_\alpha > 0$,
- *if* $\xi^*_\alpha < 0$, *then Jensen's inequality leads to*

$$V_f(\omega, \xi^*_\alpha, u) \geq \xi^*_\alpha + \frac{1}{1-\alpha}\left(u. \int x\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi^*_\alpha\right)_+ = -\frac{\alpha}{1-\alpha}\xi^*_\alpha > 0,$$

- *if* $\xi^*_\alpha = 0$, $\int v_f\left(\xi^*_\alpha, u, 0, x\right)\Pi(\omega, \mathrm{d}x, \mathrm{d}y) = \frac{1}{1-\alpha}\mathbb{E}\left[\left(u.X\right)_+ \mid \mathcal{F}\right](\omega)$. *Now, if* $\mathbb{E}\left[\left(u.X\right)_+ \mid \mathcal{F}\right] = 0$, *then* $\mathbb{E}\left[|u.X| \mid \mathcal{F}\right] = 0$, *since* $\mathbb{E}\left[u.X \mid \mathcal{F}\right] = 0$. *Then* $u.X = 0$ *a.s., so that it implies that* $u = 0$, *which is impossible.*

The right-hand sides of (5.20) and (5.22) show that the two optimization problems (5.10) and (5.11) can be written

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\theta.X\right)\right] = \mathbb{E}\left[\min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi,\theta)\right], \tag{5.23}$$

and,

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F})} \text{CVaR}_\alpha\left(L-\theta.X\right) = \inf_{\xi\in\mathbb{R}} \mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right], \tag{5.24}$$

respectively. Consequently, for all $\omega\in\Omega$, we have to solve deterministic optimization problems. Next result provides a characterization of those minima and will allow us to devise (later on) numerical procedures to estimate the quantities of interest.

Proposition 5.2.7. *Suppose Assumption 5.2.4 is satisfied. Then, for all* $\xi\in\mathbb{R}$

$$\text{Arg}\min V_f(\xi,.) = \left\{\theta\in\mathbb{R}^d |\ \nabla_\theta V_f(\xi,\theta)=0\right\} \neq \varnothing.$$

If Assumption 5.2.5 is satisfied then

$$\text{Arg}\min V_f = \left\{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d |\ \nabla_{(\xi,\theta)} V_f(\xi,\theta)=0\right\} \neq \varnothing.$$

where the gradient of V_f *can be represented for every* $(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d$ *by*

$$\nabla_{(\xi,\theta)} V_f(\xi,\theta) = \int \nabla_{(\xi,\theta)} v_f(\xi,\theta,y,x)\Pi(\mathrm{d}x,\mathrm{d}y) \tag{5.25}$$

and,

$$\nabla_\theta V_f(\xi,\theta) = \int \nabla_\theta v_f(\xi,\theta,y,x)\Pi(\mathrm{d}x,\mathrm{d}y). \tag{5.26}$$

Moreover, $\xi\mapsto\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right]$ *is Lipschitz continuous, convex,* $\lim_{|\xi|\to+\infty}\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right] = +\infty$. *Consequently,* (5.10) *and* (5.11) *admit solutions.*

Proof. Since the functions $(\xi,\theta)\mapsto v_f(\xi,\theta,y,x)$, $(y,x)\in\mathbb{R}\times\mathbb{R}^d$, are convex, the function V_f is convex. To justify the formal differentiation of V_f to get (5.25) and (5.26), we only need to check the domination property. First note that we have, for all $(y,x)\in\mathbb{R}\times\mathbb{R}^d$

$$\begin{aligned}
\frac{\partial v_f}{\partial\xi}(\xi,\theta,y,x) &= 1-\frac{1}{1-\alpha}\mathbf{1}_{\{y-\theta.x\geq\xi\}},\\
\frac{\partial v_f}{\partial\theta}(\xi,\theta,y,x) &= -\frac{1}{1-\alpha}x\mathbf{1}_{\{y-\theta.x\geq\xi\}},
\end{aligned}$$

so that there exists $C>0$ such that

$$\left|\nabla_{(\xi,\theta)} v_f(\xi,\theta,L,X)\right| \leq C\left(1+|X|\right)\in\mathbb{L}^1_{\mathbb{R}}\left(\mathbb{P}\right).$$

Now, let $\xi,\ \xi'\in\mathbb{R}$, there exists a real constant $K>0$ such that

$$\begin{aligned}
\left|\mathbb{E}\left[\inf_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right] - \mathbb{E}\left[\inf_{\theta\in\mathbb{R}^d} V_f(\xi',\theta)\right]\right| &\leq \mathbb{E}\left[\sup_{\theta\in\mathbb{R}^d}\left|V_f(\xi,\theta)-V_f(\xi',\theta)\right|\right]\\
&\leq K\ |\xi-\xi'|,
\end{aligned}$$

and, owing to Jensen's inequality

$$\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right] \geq \xi + \frac{1}{1-\alpha}\left(\mathbb{E}\left[L\right]-\xi\right)_+,$$

so that $\lim_{|\xi|\to+\infty}\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right]=+\infty$. This completes the proof.

□

5.2.4 CVaR hedging using a dynamic self financed strategy

In this section, we adress the main problem (5.1), that is hedging a contingent claim with a dynamic self-financed strategy starting with an initial investment of 0 using a static CVaR criterion at a fixed time $t=0$. Actually, in this theoretical section, we consider the more general multistage stochastic optimization problem:

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}} \mathrm{CVaR}_\alpha\left(L-\sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell\right) = \inf_{\xi\in\mathbb{R}}\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell-\xi\right)_+\right] \tag{5.27}$$

where $\mathcal{A}_{\mathcal{F}}$ is the set of all sequences $\theta=(\theta_0,\cdots,\theta_{M-1})$ such that $\theta_\ell\in L^0(\mathcal{F}_\ell)$, $\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\cdots\mathcal{F}_{M-1}\subseteq\mathcal{A}$ fixed σ-algebras. Later on, for numerical applications, we will set $\mathcal{F}_\ell$ to $\mathcal{G}_\ell$, $\ell=0,\cdots,M-1$.
Note that in order to solve (5.27), we may adress firstly the multistage stochastic optimization problem

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell-\xi\right)_+\right], \quad \text{for each } \xi\in\mathbb{R}. \tag{5.28}$$

Up to the change of variable $L:=L-\xi$, we can suppose that $\xi=0$ and $\alpha=0$ so that, without loss of generality, the problem (5.28) is equivalent to minimizing the shortfall risk

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell\right)_+\right]. \tag{5.29}$$

The optimization (5.29) is a classical stochastic control problem. One may think that it is possible to derive existence of solutions of problem (5.29) using results on dynamic programming. Unfortunately, in our case, standard assumptions of dynamic programming are not fulfilled (see e.g. [29]).

However, we can adapt this classical approach in order to derive the existence of an optimal shortfall-hedging sequence $\tilde{\theta}:=(\tilde{\theta}_\ell)_{0\leq\ell\leq M-1}$ solution of (5.29), thus we obtain the existence of an optimal CVaR-hedging sequence $\theta^*_\alpha:=(\theta^*_{\ell,\alpha})_{0\leq\ell\leq M-1}$ solution of (5.28). Finally, we will come back to (5.27) and using similar arguments to those of the static framework, we will deduce the existence of ξ^*_α solution of the problem

$$\inf_{\xi\in\mathbb{R}}\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell-\xi\right)_+\right] = \inf_{\xi\in\mathbb{R}}\mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\sum_{\ell=1}^{M}\theta^*_{\ell-1,\alpha}.\Delta X_\ell-\xi\right)_+\right].$$

In order to derive similar results to those obtained in Section 5.2.3, we consider a family of regular conditional distributions $(\Pi_\ell)_{0\leq\ell\leq M-1}$ where $\Pi_\ell(\mathrm{d}y,\mathrm{d}x) = \Pi_\ell(\omega,\mathrm{d}y,\mathrm{d}x)$ denotes the regular conditional distribution of the couple $(L,\Delta X_1,\cdots,\Delta X_M)$ given $\mathcal{F}_\ell$ and we make the following assumption.

Assumption 5.2.8.

i) *The distribution of* $(L,\Delta X_1,\cdots,\Delta X_M)$ *satisfies* $L\in\mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $\Delta X_\ell\in\mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$, $\ell=1,\cdots,M$

ii) $\operatorname{ess\,inf}_{u\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1}\mathbb{E}\left[(u.\Delta X_\ell)_+\mid\mathcal{F}_{\ell-1}\right]>0$ *a.s.*

Remark 5. *In the same way that Remark 4 in the one step framework, the non-degeneracy Assumption 5.2.8 can be replaced by the stronger assumption*

- $\mathbb{E}\left[X_\ell\mid\mathcal{F}_{\ell-1}\right]=0$ *and* $\mathbb{E}\left[X_\ell X_\ell^T\mid\mathcal{F}_{\ell-1}\right]$ *is a.s. positive definite in* $\mathcal{S}(d,\mathbb{R})$ *for* $\ell=1,\cdots,M$.
- *The conditional distribution of* X_ℓ *given* $\mathcal{F}_{\ell-1}$ *is continuous (no affine hyperplane has positive mass) for* $\ell=1,\cdots,M$.

In the spirit of the dynamic programming principle, we construct the solution of (5.29) using a step by step backward induction. To be more precise, using similar arguments to those used to prove (5.21), one first notices that (5.29) can be written

$$\inf_{\theta_\ell\in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell,\mathbb{P}),\ell=0,\cdots,M-2}\mathbb{E}\left[\operatorname*{ess\,inf}_{\theta_{M-1}\in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1},\mathbb{P})}\mathbb{E}\left[\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell\right)_+\middle|\mathcal{F}_{M-1}\right]\right], \tag{5.30}$$

so that one may start by solving the following problem

$$\operatorname*{ess\,inf}_{\theta_{M-1}\in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1},\mathbb{P})}\mathbb{E}\left[\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell\right)_+\middle|\mathcal{F}_{M-1}\right](\omega)=\min_{\theta_{M-1}\in\mathbb{R}^d}V_{M-1}(\omega,\theta_{0:M-2},\theta_{M-1}) \tag{5.31}$$

$$=V_{M-1}(\omega,\theta_{0:M-2},\tilde{\theta}_{M-1})\ \ a.s. \tag{5.32}$$

where $\tilde{\theta}_{M-1}\in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1})$, V_{M-1} is defined for all $\omega\in\Omega$, $\theta_\ell\in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\ell=1,\cdots,M-1$, by

$$\begin{aligned}V_{M-1}(\omega,\theta_{0:M-2},\theta_{M-1}) &:= \mathbb{E}\left[\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell\right)_+\ \middle|\ \mathcal{F}_{M-1}\right](\omega)\\ &=\int\left(y-\sum_{\ell=1}^M\theta_{\ell-1}(\omega).\Delta x_\ell\right)_+\Pi_{M-1}(\omega,\mathrm{d}x,\mathrm{d}y).\end{aligned} \tag{5.33}$$

This follows from similar arguments to those of the proof of Proposition 5.2.5, *i.e.* from the fact that for all $\omega\in\Omega$, $\theta_\ell\in L^0(\mathcal{F}_\ell)$, $\ell=1,\cdots,M-2$, the function (defined on $\mathbb{R}^d$) by $\theta_{M-1}\mapsto V_{M-1}(\xi,\theta_{0:M-2},\theta_{M-1})$ is convex, Lipschitz continuous and

$\lim_{|\theta_{M-1}|\to+\infty} V_{M-1}(\xi,\theta_{0:M-2},\theta_{M-1}) = +\infty$. Thus, it implies that (5.31) has a solution that we denote by $\tilde{\theta}_{M-1} := \tilde{\theta}_{M-1}(\omega,\theta_0,\cdots,\theta_{M-2})$, which is $\mathcal{F}_{M-1}$-measurable (owing to measurable selection, see e.g. Lemma 3 and Lemma 4 in [29]) so that (5.32) holds.

Then we proceed by a backward induction: we denote by $\tilde{\theta}_{\ell:M-1} := (\tilde{\theta}_\ell,\cdots,\tilde{\theta}_{M-1})$ the solution built down to step ℓ. At step $\ell-1$, we adress for every $\theta_{0:\ell-2} \in L^0_{\mathbb{R}^d}(\mathcal{F}_0,\mathbb{P}) \times \cdots \times L^0_{\mathbb{R}^d}(\mathcal{F}_{\ell-2},\mathbb{P})$, the problem

$$\begin{aligned}\operatorname*{ess\,inf}_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{F}_{\ell-1},\mathbb{P})} \mathbb{E}\left[V_\ell(\theta_{0:\ell-1},\tilde{\theta}_\ell)\middle|\mathcal{F}_{\ell-1}\right](\omega) &= \min_{\theta_{\ell-1}\in\mathbb{R}^d} V_{\ell-1}(\omega,\theta_{0:\ell-2},\theta_{\ell-1}) \\ &= V_{\ell-1}(\omega,\theta_{0:\ell-2},\tilde{\theta}_{\ell-1}) \ \ a.s. \end{aligned} \tag{5.34}$$

where for all $\theta_k \in L^0_{\mathbb{R}^d}(\mathcal{F}_k,\mathbb{P})$, $k=0,\cdots,\ell-1$, the functions V_ℓ and $V_{\ell-1}$ are defined by

$$\begin{aligned} V_\ell(\omega,\theta_{0:\ell-1},\tilde{\theta}_\ell) &:= \mathbb{E}\left[\left(L-\sum_{k=1}^{\ell}\theta_{k-1}.\Delta X_k - \sum_{k=\ell+1}^{M}\tilde{\theta}_{k-1}.\Delta X_k\right)_+ \middle|\ \mathcal{F}_\ell\right](\omega) \\ &= \int\left(y-\sum_{k=1}^{\ell}\theta_{k-1}(\omega).\Delta x_k - \sum_{k=\ell+1}^{M}\tilde{\theta}_{k-1}.\Delta x_k\right)_+ \Pi_\ell(\omega,\mathrm{d}x,\mathrm{d}y) \end{aligned} \tag{5.35}$$

and,

$$V_{\ell-1}(\omega,\theta_{0:\ell-2},\theta_{\ell-1}) = \int\left(y-\sum_{k=1}^{\ell}\theta_{k-1}(\omega).\Delta x_k - \sum_{k=\ell+1}^{M}\tilde{\theta}_{k-1}.\Delta x_k\right)_+ \Pi_{\ell-1}(\omega,\mathrm{d}x,\mathrm{d}y) \tag{5.36}$$

The following proposition implies that (5.27) has an optimal solution $(\xi^*_\alpha,\theta^*_\alpha) \in \mathbb{R}\times\mathcal{A}_\mathcal{F}$.

Proposition 5.2.9. *Suppose that Assumtpion 5.2.8 is satisfied. Then,*

i) (5.30) *is satisfied, the problem* (5.31) *has a solution and for* $\ell = M-1,\cdots,1$, *one can find an* $\mathcal{F}_{\ell-1}$*-measurable random variable* $\tilde{\theta}_{\ell-1}$ *solution of* (5.34). *Thus,* (5.28) *has an optimal solution denoted by* $\theta^*_\alpha := (\theta^*_{\alpha,\ell})_{0\le\ell\le M-1}$.

ii) The function $\xi \mapsto \xi + \frac{1}{1-\alpha}\inf_{\theta\in\mathcal{A}_\mathcal{F}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell - \xi\right)_+\right]$ *is convex, Lipschitz continuous and satisfies* $\lim_{|\xi|\to+\infty}\xi+\frac{1}{1-\alpha}\inf_{\theta\in\mathcal{A}_\mathcal{F}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell-\xi\right)_+\right] = +\infty$ *so that* (5.27) *admits a solution.*

Proof. i) The proof of (5.30) uses similar arguments than those used in the proof of (5.19).
Let $\tilde{\theta}_{M-1}$ be a solution of (5.31). We go one step backward. For all $\omega\in\Omega$, $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\ell=0,\cdots,M-3$, (using the definition of V_{M-1}) we are interested by the

function

$$\theta_{M-2} \mapsto V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2}) :=$$

$$\mathbb{E}\left[\underbrace{\operatorname*{ess\,inf}_{\theta_{M-1}\in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M-2}\theta_{\ell-1}.\Delta X_\ell - \theta_{M-2}.\Delta X_{M-1} - \tilde{\theta}_{M-1}.\Delta X_M\right)_+ \middle| \mathcal{F}_{M-1}\right]}_{=V_{M-1}(\theta_{0:M-2},\tilde{\theta}_{M-1})} \middle| \mathcal{F}_{M-2}\right](\omega).$$

It is straightforward that this function is convex. Let $\theta_{M-2},\ \theta'_{M-2} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-2})$, using the standard inequality $|\operatorname{ess\,inf}_{i\in I} a_i - \operatorname{ess\,inf}_{i\in I} b_i| \leq \operatorname{ess\,sup}_{i\in I} |a_i - b_i|$, we have

$$\left|V_{M-2}(\theta_{0:M-3}, \theta_{M-2}) - V_{M-2}(\theta_{0:M-3}, \theta'_{M-2})\right| \leq |\theta_{M-2} - \theta'_{M-2}|\mathbb{E}\left[|\Delta X_{M-1}| \mid \mathcal{F}_{M-2}\right] \ a.s.$$

so that the function is Lipschitz continuous.

Lemma 5.2.10. *(Conditional Jensen's inequaility) Let F be a non-negative convex function and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. If X is random variable such that $\mathbb{E}[|X| \mid \mathcal{B}] < +\infty$, a.s. then*

$$\mathbb{E}[F(X) \mid \mathcal{B}] \geq F(\mathbb{E}[X \mid \mathcal{B}]) \quad a.s.$$

Proof. This is a straightforward adaptation of the proof of Jensen's inequality. □

Now, owing to Lemma 5.2.9 , we have

$$V_{M-2}(\theta_{0:M-3}, \theta_{M-2}) \geq \mathbb{E}\left[\left(\mathbb{E}[L \mid \mathcal{F}_{M-1}] - \sum_{\ell=1}^{M-2}\theta_{\ell-1}.\Delta X_\ell - \theta_{M-2}.\Delta X_{M-1}\right)_+ \middle| \mathcal{F}_{M-2}\right]. \tag{5.37}$$

We aim at showing that the right-hand side of (5.37) goes to infinity as $|\theta_{M-2}| \to +\infty$. First, the sub-additivity of the function $x \mapsto x_+$ implies that

$$\begin{aligned}\mathbb{E}\left[\ (-\theta_{M-2}.\Delta X_{M-1})_+ \middle| \mathcal{F}_{M-2}\right] \leq\ & \mathbb{E}\left[\left(\mathbb{E}[L|\mathcal{F}_{M-1}] - \sum_{\ell=1}^{M-1}\theta_{\ell-1}.\Delta X_\ell\right)_+ \middle| \mathcal{F}_{M-2}\right] \\ & + \mathbb{E}\left[\left(-\mathbb{E}[L|\mathcal{F}_{M-1}] + \sum_{\ell=1}^{M-2}\theta_{\ell-1}.\Delta X_\ell\right)_+ \middle| \mathcal{F}_{M-2}\right].\end{aligned}$$

We focus on the left-hand side of the above inequality, this quantity is lower bounded by

$$|\theta_{M-2}| \inf_{u\in\mathcal{S}_d(0,1)} \mathbb{E}\left[(u.\Delta X_{M-1})_+ \middle| \mathcal{F}_{M-2}\right],$$

where $\mathcal{S}_d(0,1) := \left\{u \in \mathbb{R}^d \mid |x| = 1\right\}$ denotes the (compact) unit sphere. Using Assumtpion 5.2.7 and similar arguments than those of the proof of Proposition 5.2.5, one shows that

$$\inf_{u\in\mathcal{S}_d(0,1)} \mathbb{E}\left[(u.\Delta X_{M-1})_+ \middle| \mathcal{F}_{M-2}\right] > 0, \quad a.s.$$

Thus, for all $\omega \in \Omega$, $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\ell = 0, \cdots, M-3$, $\lim_{|\theta_{M-2}|\to+\infty} V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2}) = +\infty$ and the function $\theta_{M-2} \mapsto V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2})$ has a minimum $\tilde{\theta}_{M-2}$ which is $\mathcal{F}_{M-2}$-measurable owing to measurable selection theorem.

Furthermore, using similar arguments to those used for the proof of Proposition 5.2.5, one shows that for all $\omega \in \Omega$, for every $\theta_{0:M-3} \in L^0_{\mathbb{R}^d}(\mathcal{F}_0) \times \cdots \times L^0_{\mathbb{R}^d}(\mathcal{F}_{M-3})$

$$\operatorname*{ess\,inf}_{\theta_{M-2} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-2})} V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2}) =$$

$$\operatorname*{ess\,inf}_{(\theta_{M-2},\theta_{M-1}) \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-2}) \times L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell \right)_+ \mid \mathcal{F}_{M-2} \right] \quad a.s.$$

Using similar arguments, one shows that if the solution is built down to step ℓ, for all $\omega \in \Omega$ and $\theta_k \in L^0(\mathcal{F}_k)$, $k = 0, \cdots, \ell-2$, the function $\theta_{\ell-1} \mapsto V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1})$ is convex, Lipschitz continuous and satisfies $\lim_{|\theta_{\ell-1}|\to+\infty} V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1}) = +\infty$. Consequently, there exists $\tilde{\theta}_{\ell-1}$ solution of (5.34). Thus, (5.28) has an optimal $\theta^*_\alpha := (\theta^*_{\alpha,\ell})_{0\le\ell\le M-1}$.

Now we come back to our original problem

$$\inf_{\xi\in\mathbb{R}} \xi + \frac{1}{1-\alpha} \inf_{\theta\in\mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi \right)_+ \right] =$$

$$\inf_{\xi\in\mathbb{R}} \xi + \frac{1}{1-\alpha} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta^*_{\alpha,\ell-1}.\Delta X_\ell - \xi \right)_+ \right].$$

For all $x \in \mathbb{R}$, the functions $\xi \mapsto \xi + \frac{1}{1-\alpha}(x-\xi)_+$ are convex and Lipschitz continuous so that $\xi \mapsto \xi + \frac{1}{1-\alpha} \inf_{\theta\in\mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi \right)_+ \right]$ is convex, Lipschitz continuous. Owing to Assumtpion 5.2.7 and using Lemma 5.2.9 with a backward induction, one shows

$$\mathbb{E}\left[\xi + \frac{1}{1-\alpha} \left(L - \sum_{\ell=1}^{M} \theta^*_{\alpha,\ell-1}.\Delta X_\ell - \xi \right)_+ \right] \ge \xi + \frac{1}{1-\alpha} \left(\mathbb{E}[L] - \xi \right)_+ ,$$

so that $\lim_{|\xi|\to+\infty} \xi + \frac{1}{1-\alpha} \inf_{\theta\in\mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi \right)_+ \right] = +\infty$. This completes the proof. □

5.3 Computational and numerical aspects of CVaR hedging

In this section, we propose several methods to compute the optimal strategies of the three problems (5.3), (5.2) and (5.1). First, we will focus on (5.2) since it will be the main building block when we are going to propose several algorithms to approximate the optimal dynamic strategy solution of (5.1).

5.3.1 Markovian framework and Optimal Vector Quantization

In order to simplify the numerical computation of conditional expectations that appear in the problem (5.2), we will work under Assumptions 5.1.1 and 5.1.2.

To be more precise, from a modeling point of view, it is quite natural to consider that the random variable L can be written as a function of the process (X, Z), *i.e.* $L = \phi(X, Z)$. Typically, in the electricity market, Z can be considered as the temperature process and may influence electricity spot prices and electricity forward prices. Consequently, we assume that there exists two continuous functions $F : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}$ and $G : \left(\mathbb{R}^d\right)^{\ell_0+1} \times \left(\mathbb{R}^q\right)^{\ell_0+1} \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}^d$ such that

$$X_M - X_{\ell_0} = G(X_{\ell_0}, Z_{\ell_0}, U_{\ell_0} + 1) \text{ and } L = F(X_{\ell_0}, Z_{\ell_0}, U_{\ell_0+1})$$

where U_{ℓ_0+1} is a r_{ℓ_0}-dimensional random variable independent of $\mathcal{G}_{\ell_0} := \sigma(X_{\ell_0}, Z_{\ell_0})$. We will denote U for U_{ℓ_0+1}. Under this markovian framework, the function (5.15) can be written for all $(x, z) \in \mathbb{R}^d \times \mathbb{R}^q$

$$V(\xi, \theta, x, z) = \mathbb{E}\left[v(\xi, \theta, x, z, U)\right],$$

where $v(\xi, \theta, x, z, u) := \xi + \frac{1}{1-\alpha}\left(F(x, z, u) - \theta.G(x, z, u) - \xi\right)_+$ so that (5.22) becomes

$$\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0}, \mathbb{P}), \xi \in L^0_{\mathbb{R}}(\mathcal{G}_{\ell_0}, \mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ \middle| \mathcal{G}_{\ell_0}\right] = \left(\min_{(\xi,\theta) \in \mathbb{R} \times \mathbb{R}^d} V(\xi, \theta, x, z)\right)_{(x,z)=(X_{\ell_0}, Z_{\ell_0})} \quad a.s. \tag{5.38}$$

Consequently, in order to solve the global problem (5.2) we need to solve the local optimization problem that appears in the right-hand side of the above equation for each $(X_{\ell_0}(\omega), Z_{\ell_0}(\omega))$. Then, we have to estimate the quantity

$$\mathbb{E}\left[\left(\inf_{(\theta,\xi) \in \mathbb{R}^d \times \mathbb{R}} V(\xi, \theta, x, z)\right)_{|x=X_{\ell_0}, z=Z_{\ell_0}}\right]. \tag{5.39}$$

When the dimension of the random variable (X_{ℓ_0}, Z_{ℓ_0}) is large (greater than $5, 10$), one can use Monte-Carlo simulations and estimates (5.39) using N samples by

$$\frac{1}{N}\sum_{k=1}^{N}\left(\inf_{(\theta,\xi) \in \mathbb{R}^d \times \mathbb{R}} V(\xi, \theta, x, z)\right)_{|x=X_{\ell_0,k}, z=Z_{\ell_0,k}},$$

where $\left(X_{\ell_0,k}, Z_{\ell_0,k}\right)_{1 \le k \le N}$ are i.i.d. random vectors having the distribution of (X_{ℓ_0}, Z_{ℓ_0}).

When the dimension of the random variable (X_{ℓ_0}, Z_{ℓ_0}) is small (say less than $5, 10$), we can use an integration cubature formula based for instance on a spatial discretization of (X_{ℓ_0}, Z_{ℓ_0}). A commonly used method in such a framework is optimal vector quantization. Thus, we consider an optimal N_{ℓ_0}-quantization $\left(\hat{X}_{\ell_0}, \hat{Z}_{\ell_0}\right)$ of the random variable (X_{ℓ_0}, Z_{ℓ_0}), based on an optimal quantization grid $\Gamma_{\ell_0} := \Gamma^{N_{\ell_0}}_{(X_{\ell_0}, Z_{\ell_0})} = \left(\left(x^1_{\ell_0}, z^1_{\ell_0}\right), \cdots, \left(x^{N_{\ell_0}}_{\ell_0}, z^{N_{\ell_0}}_{\ell_0}\right)\right)$. Then if we denote $CV^*_\alpha(x^j_{\ell_0}, z^j_{\ell_0})$

for $\inf_{(\theta,\xi)\in\mathbb{R}^d\times\mathbb{R}} V(\xi,\theta,(x_{\ell_0}^j,z_{\ell_0}^j))$, $j=1,\cdots,N_{\ell_0}$, the quantization based quadrature formula to approximate (5.39) is given by

$$\sum_{j=1}^{N_{\ell_0}} CV_\alpha^*(x_{\ell_0}^j,z_{\ell_0}^j)\mathbb{P}\left((X_{\ell_0},Z_{\ell_0})\in C_j(x_{\ell_0},z_{\ell_0})\right), \tag{5.40}$$

where $(C_j(x_{\ell_0},z_{\ell_0}))_{j=1,\cdots,N_{\ell_0}}$ is a Voronoi tesselation of the N_{ℓ_0}-quantizer Γ_{ℓ_0}. Fore more details about optimal vector quantization, including error bounds for cubature formulae, we refer to [72].

Consequently, we need to compute the solution as well as the value of the objective function for all nodes of a quantification grid (or for all Monte-Carlo samples). Thus, throughout this section, we will focus on the value function that appears within the brackets of the right-hand side of (5.38).

For the sake of simplicity, we will temporarily drop (x,z) in the notations so that we will denote $F(U)$ for $F(x,z,U)$, $G(U)$ for $G(x,z,U)$, $V(\xi,\theta)$ for $V(\xi,\theta,x,z)$, $V(\xi,\theta,U)$ for $v(\xi,\theta,x,z,U)$ and so on. Thus, we will omit "for all $(x,z)\in\mathbb{R}^d\times\mathbb{R}^q$" in any assumption or property given below about those distributions or functions.

5.3.2 Computing CVaR hedging by stochastic approximation: a first approach

The above local representation (5.38) naturally yields a stochastic gradient algorithm derived from the Lyapunov function V which will converge toward $(\xi_\alpha^*,\theta_\alpha^*)\in \operatorname{Arg\,min} V$. Then, following the procedure investigated in [6], a companion recursive procedure can be easily devised which has $CV_\alpha^* := CV_\alpha^*(x,z) = \mathrm{CVaR}_\alpha\left(F(U)-\theta_\alpha^*.G(U)\right)$ as a target, *i.e.*, the CVaR_α of the CVaR-hedged portfolio at the point (x,z). Finally, in order to compute the value function at time t_ℓ *i.e.* the global expectation (5.39), we will rely on the cubature formula based on optimal vector quantization given by (5.40).

First we set,

$$H_1(\xi,\theta,U) := \frac{\partial v}{\partial\xi}(\xi,\theta,U) = 1-\frac{1}{1-\alpha}\mathbf{1}_{\{F(U)-\theta.G(U)\geq\xi\}}, \tag{5.41}$$

$$H_{2:d+1}(\xi,\theta,U) := \frac{\partial v}{\partial\theta}(\xi,\theta,U) = -\frac{1}{1-\alpha}G(U)\mathbf{1}_{\{F(U)-\theta.G(U)\geq\xi\}}, \tag{5.42}$$

so that,

$$\nabla_{(\xi,\theta)}V(\xi,\theta) = \mathbb{E}\left[\left(H_1(\xi,\theta,U),H_{2:d+1}(\xi,\theta,U)\right)\right].$$

Since we are looking for (ξ,θ) for which $\mathbb{E}\left[H_1(\xi,\theta,U)\right]$ and $\mathbb{E}\left[H_{2:d+1}(\xi,\theta,U)\right]=0$, we implement a classical R.M. algorithm to approximate $(\xi_\alpha^*,\theta_\alpha^*)$, *i.e.*, we define recursively for $n\geq 1$:

$$\xi_n = \xi_{n-1}-\gamma_n H_1(\xi_{n-1},\theta_{n-1},U_n), \tag{5.43}$$

$$\theta_n = \theta_{n-1}-\gamma_n H_{2:d+1}(\xi_{n-1},\theta_{n-1},U_n), \tag{5.44}$$

where $(U_n)_{n\geq1}$ is an i.i.d. sequence of random vectors with the same distribution as U, independent of (ξ_0,θ_0), with $\xi_0\in L^2_{\mathbb{R}}(\mathbb{P})$, $\theta_0\in L^2_{\mathbb{R}^d}(\mathbb{P})$ and $(\gamma_n)_{n\geq1}$ is a positive

deterministic step sequence satisfying

$$\sum_{n\geq 1}\gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1}\gamma_n^2 < +\infty. \tag{5.45}$$

Following [6], as a second step, in order to estimate $\mathrm{CVaR}_\alpha\left(F(U) - \theta_\alpha^*.G(U)\right)$, *i.e.* the CVaR_α of the local CVaR hedged loss, we devise a *companion* procedure using the same step sequence than (5.43) and (5.44), for $n \geq 1$

$$C_n = C_{n-1} - \gamma_n H_{d+2}\left(\xi_{n-1}, \theta_{n-1}, C_{n-1}, U_n\right), \tag{5.46}$$

with $H_{d+2}\left(\xi, \theta, c, u\right) := c - v(\xi, \theta, u)$. In order to derive the *a.s.* convergence of (5.43), (5.44) and (5.46), we introduce the following additional assumption on the distribution of $F(U)$ and $G(U)$.

Assumption 5.3.1. *Let $a > 0$. $F(U) \in \mathbb{L}^{2a}(\mathbb{P})$ and $G(U) \in \mathbb{L}^{2a}(\mathbb{P})$.*

To establish the *a.s.* convergence of $(\xi_n, \theta_n, C_n)_{n\geq 1}$, we will rely on Robbins-Monro Theorem (see e.g. [23]). In fact we will use the following slight extension (which takes into acount the case of non-uniqueness of the target). For a proof, we refer e.g. to [62].

Theorem 5.3.2. *(Extended Robbins-Monro Theorem) Let $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function and let X be an $\mathbb{R}^d$-valued random vector such that $\mathbb{E}[|H(y, X)|] < +\infty$ for every $y \in \mathbb{R}^d$. Then set*

$$\forall y \in \mathbb{R}^d, \quad h(y) = \mathbb{E}[H(y, X)].$$

Suppose that the function h is continuous and that $\mathcal{T}^* := \{h = 0\}$ satisfies

$$\forall y \in \mathbb{R}^d \ \ |\mathcal{T}^*, \forall y^* \in \mathcal{T}^*, \quad \langle y - y^*, h(y)\rangle > 0. \tag{5.47}$$

Let $(\gamma_n)_{n\geq 1}$ be a deterministic step sequence satisfying (5.45). Suppose that

$$\forall y \in \mathbb{R}^d, \quad \mathbb{E}[|H(y, X)|^2] \leq C(1 + |y|^2) \tag{5.48}$$

(which implies that $|h(y)|^2 \leq C(1 + |y|)$).
Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence of random vectors having the distribution of X, let y_0 be a random vector independent of $(X_n)_{n\geq 1}$ satisfying $\mathbb{E}|y_0|^2 < +\infty$, all defined on the same probability space $(\Omega, A, \mathbb{P})$.
Then, the recursive procedure defined for $n \geq 1$ by

$$y_n = y_{n-1} - \gamma_n H(y_{n-1}, X_n),$$

satisfies:

$$\exists\, y_\infty : (\Omega, \mathcal{A}) \to \mathcal{T}^*, y_\infty \in \mathbb{L}^2\left(\mathbb{P}\right) \quad \textit{such that} \quad y_n \xrightarrow{a.s.} y_\infty.$$

The convergence also holds in $L^p(\mathbb{P}), p \in (0, 2)$.

In the next proposition, we establish the *a.s.* convergence of the sequence $(\xi_n, \theta_n, C_n)_{n\geq 1}$ toward its target $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$.

Theorem 5.3.3. *Suppose that Assumptions 5.2.4 and 5.3.1 are satisfied (for $a = 1$), and that the step sequence $(\gamma_n)_{n\geq 1}$ satisfies the usual decreasing step assumption* (5.45).
Then the recursive procedure defined by (5.43), (5.44) *and* (5.46) *satisfies:*

$$\exists\ (\xi_\alpha^*, \theta_\alpha^*) : (\Omega, \mathcal{A}) \to \operatorname{Arg\,min} V, \textit{ (which is a compact set)},$$

such that

$$(\xi_n, \theta_n) \xrightarrow{a.s.} (\xi_\alpha^*, \theta_\alpha^*), \quad n \to +\infty.$$

Moreover,

$$C_n \xrightarrow{a.s.} CV_\alpha^* = \min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V(\xi, \theta) = V(\xi_\alpha^*, \theta_\alpha^*) \quad n \to +\infty.$$

Proof. We first prove the *a.s.* convergence of $(\xi_n, \theta_n)_{n\geq 1}$ using the above Extended Robbins-Monro Theorem; that of $(C_n)_{n\geq 1}$ will follow the lines of the proof of the *a.s.* convergence of the CVaR algorithm in [6] (Section 2.2). In order to apply the R.M. theorem, we have to check the following facts:

- *Mean reversion.* For the sake of simplicity, we denote by y the couple (ξ, θ). The mean function of the algorithm defined by (5.43) and (5.44) reads

  $$l(y) := \mathbb{E}\left[\left(H_1\left(y, U\right), H_{2:d+1}\left(y, U\right)\right)\right] = \nabla V\left(y\right)$$

  so that $\mathcal{T}^* := \{l = 0\} = \{\nabla V = 0\}$. Moreover, if $y^* \in \mathcal{T}^*$ and $y \in \mathbb{R}\times\mathbb{R}^d \backslash \mathcal{T}^*$,

  $$\langle y - y^*, l(y)\rangle = \langle y - y^*, \nabla V\left(y\right)\rangle > 0,$$

  since the function V is a convex differentiable function and $\operatorname{Arg\,min} V$ is non empty.

- *Linear Growth of* $(\xi, \theta) \mapsto \mathbb{E}\left[\left|H_1\left(\xi, \theta, U\right)\right|^2 + \left|H_{2:d+1}\left(\xi, \theta, U\right)\right|^2\right]$. This conditions is clearly fulfilled since there exists a real constant $C > 0$ such that

  $$\mathbb{E}\left[\left|H_1\left(\xi, \theta, U\right)\right|^2\right] < C \text{ and } \mathbb{E}\left[\left|H_{2:d+1}\left(\xi, \theta, U\right)\right|^2\right] < \frac{1}{\left(1-\alpha\right)^2}\mathbb{E}\left[\left|G(U))\right|^2\right] < C,$$

  so that,

  $$\mathbb{E}\left[\left|H_1\left(\xi, \theta, U\right)\right|^2 + \left|H_{2:d+1}\left(\xi, \theta, U\right)\right|^2\right] \leq C\left(1 + |y|^2\right).$$

Consequently, we have

$$(\xi_n, \theta_n) \xrightarrow{a.s.} (\xi_\alpha^*, \theta_\alpha^*).$$

In order to prove the *a.s.* convergence of $(C_n)_{n\geq 1}$ toward CV_α^*, we set for convenience $\gamma_0 := 1 + \sup_{n\geq 1}\gamma_n$. Then, one defines recursively a sequence $(\Delta_n)_{n\geq 1}$ by

$$\Delta_{n+1} = \Delta_n \frac{\gamma_{n+1}}{\gamma_n}\frac{\gamma_0}{\gamma_0 - \gamma_{n+1}},\ n \geq 0,\ \Delta_0 = 1.$$

Elementary computations show by induction that

$$\gamma_n = \gamma_0 \frac{\Delta_n}{S_n},\ n \geq 0,\ \text{with } S_n := \sum_{k=0}^{n}\Delta_k. \tag{5.49}$$

Furthermore, it follows from (5.49) that for every $n \geq 1$, $\log(S_n) \geq \frac{1}{\gamma_0}\sum_{k=1}^{n}\gamma_k$, which implies that $\lim_n S_n = +\infty$.
Now using (5.46) and (5.49), one gets for every $n \geq 1$

$$S_n C_n = S_{n-1}C_{n-1} + \Delta_n \left(\Delta N_n + V(\xi_{n-1}, \theta_{n-1})\right)$$

where, $\Delta N_n := v(\xi_{n-1}, \theta_{n-1}, U_n) - V(\xi_{n-1}, \theta_{n-1})$, $n \geq 1$, defines a sequence of martingale increments with respect to the natural filtration of the algorithm $\mathcal{F}_n := \sigma(\xi_0, \theta_0, U_1, \cdots, U_n),\ n \geq 0$. This implies that

$$C_n = \frac{1}{S_n}\left(\sum_{k=0}^{n-1}\Delta_{k+1}\Delta N_{k+1}\right) + \frac{1}{S_n}\left(\sum_{k=0}^{n-1}\Delta_{k+1}V(\xi_k, \theta_k)\right).$$

The second term in the right hand side of the above equality converges to $V(\xi_\alpha^*, \theta_\alpha^*) = C_\alpha^*$ owing to the continuity of V at $(\xi_\alpha^*, \theta_\alpha^*)$ and Cesàro's Lemma. The convergence to 0 of the first term will follow from the *a.s.* convergence of the series

$$N_n^\gamma := \sum_{k=1}^{n}\gamma_k \Delta N_k, \quad n \geq 1,$$

by the Kronecker Lemma since $\gamma_n = \gamma_0 \frac{\Delta_n}{S_n}$. The sequence $(N_n^\gamma)_{n\geq 1}$ is an $\mathcal{F}_n$-martingale since the ΔN_k's are martingale increments and

$$\mathbb{E}\left[(\Delta N_n)^2 | \mathcal{F}_{n-1}\right] \leq \frac{1}{(1-\alpha)^2}\mathbb{E}\left[\left(F(U) - \theta.G(U)\right) - \xi\right)^2\Big]_{|\xi=\xi_{n-1}, \theta=\theta_{n-1}}.$$

Assumption 2 and the *a.s.* convergence of (ξ_n, θ_n) toward $(\xi_\alpha^*, \theta_\alpha^*)$ imply that

$$\sup_{n\geq 1}\mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < +\infty \quad a.s.$$

Consequently, the step assumption (5.45) implies $\langle N^\gamma \rangle_\infty = \sum_{n\geq 1}\gamma_n^2 \mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < \infty$, which in term yields the *a.s.* convergence of $(N_n^\gamma)_{n\geq 1}$, so that $C_n \xrightarrow{a.s.} CV_\alpha^*$. □

As concerns the rate of convergence, the global procedure composed by (5.43), (5.44), (5.46) is a regular stochastic algorithm that behaves as described in usual Stochastic Approximation textbooks like [13], [22], [57]. As soon as $\mathcal{T}^*$ is reduced to a single point $(\xi_\alpha^*, \theta_\alpha^*)$ (the local CVaR CV_α^* is always unique), the procedure satisfies under quite standard assumptions a CLT at rate $\gamma_n^{-\frac{1}{2}}$. It is well known that the best asymptotic rate is obtained by specifying $\gamma_n = \frac{c}{b+n}$, $c, b > 0$. However, the choice of c is subject to a stringent condition depending on $(\xi_\alpha^*, \theta_\alpha^*)$ (which is unknown to the user). This always induces a more or less (sub-optimal) blind choice for the constant c.

To overcome this classical problem, we introduce the empirical mean of the global algorithm implemented with a slowly decreasing step "à la Ruppert & Polyak" (see e.g. [74]). First, we write the global algorithm in a more synthetic way by setting for $n \geq 1$

$$\phi_n = (\xi_n, \theta_n, C_n), \quad \phi_0 = (\xi_0, \theta_0, C_0)$$

and

$$\phi_n = \phi_{n-1} - \gamma_n H(\phi_{n-1}, U_n), \tag{5.50}$$

where $H(\phi,u) = (H_1(\xi,\theta,u), H_{2:d+1}(\xi,\theta,u), H_{d+2}(\xi,\theta,c,u))$. Thus, the Cesàro mean of the procedure

$$\overline{\phi}_n = \frac{\phi_0 + \cdots + \phi_{n-1}}{n}, \quad n \geq 1, \tag{5.51}$$

where ϕ_n is defined by (5.50), *a.s.* converges to the same target. The Ruppert & Polyak's Averaging Principle says that an appropriate choice of the step yields for free the optimal asymptotic rate and the smallest possible asymptotic variance. We recall below this result following a version established in [73].

Theorem 5.3.4. *(Ruppert and Polyak's Averaging Principle) Suppose that the $\mathbb{R}^d$-sequence $(\phi_n)_{n\geq 0}$ is defined recursively by*

$$\phi_n = \phi_{n-1} - \gamma_n \left(h(\phi_{n-1}) + \epsilon_n\right),$$

where h is a Borel function. Let $\mathcal{F}_n := \sigma(\xi_0, \theta_0, U_1, \cdots, U_n)$ be the natural filtration of the algorithm. Suppose that h is $\mathcal{C}^1$ in the neighborhood of a zero ϕ^* of h and that $P = Dh(\phi^*)$ is a uniformly repulsive matrix (all its eigenvalues have positive real parts), and that $(\epsilon_n)_{n\geq 1}$ is a random $\mathcal{F}_n$-adapted sequence satisfying

$$(RP) \equiv \exists\, C > 0, \text{ such that a.s. } \begin{cases} (i)\ \mathbb{E}[\epsilon_{n+1}|\mathcal{F}_n]\mathbf{1}_{\{|\phi_n - \phi^*| \leq C\}} = 0, \\ (ii)\ \exists\, b > 2,\ \sup_n \mathbb{E}[|\epsilon_{n+1}|^b|\mathcal{F}_n]\ \mathbf{1}_{\{|\phi_n-\phi^*|\leq C\}} < +\infty, \\ (iii)\ \exists\, \Gamma \in \mathcal{S}^+(d,\mathbb{R}) \text{ such that } \mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right] \xrightarrow{a.s.} \Gamma. \end{cases} \tag{5.52}$$

Set $\gamma_n = \frac{\gamma_1}{n^\beta}$ with $\frac{1}{2} < \beta < 1$, and

$$\overline{\phi}_n := \frac{\phi_0 + \cdots + \phi_{n-1}}{n} = \overline{\phi}_{n-1} - \frac{1}{n}\left(\overline{\phi}_{n-1} - \phi_{n-1}\right),\ n \geq 1.$$

Then, on the set of convergence $\{\phi_n \to \phi^*\}$:

$$\sqrt{n}\left(\overline{\phi}_n - \phi^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, P^{-1}\Gamma(P^{-1})^T\right) \qquad \text{as } n \to +\infty,$$

In order to derive the convergence rate of the averaged algorithm, we suppose that the conditional distribution of $(X_M - X_{\ell_0}, L)$ given $(X_{\ell_0}, Z_{\ell_0}) = (x_{\ell_0}, z_{\ell_0})$ has a probability density function $p^{(X_{\ell_0},Z_{\ell_0})=(x_{\ell_0},z_{\ell_0})}_{(X_M - X_{\ell_0},L)}$ for all $(x_{\ell_0}, z_{\ell_0}) \in \mathbb{R}^d \times \mathbb{R}^q$ that we will denote $p_{X,L}$ for the sake of simplicity. Moreover, in order to simplify the notations, we will denote X the conditional distribution of $(X_M - X_{\ell_0})$ given $(X_{\ell_0}, Z_{\ell_0}) = (x_{\ell_0}, z_{\ell_0})$. Finally, we will denote by p_X and p_L the (conditional) marginal density functions of $X_M - X_{\ell_0}$ and L (given $(X_{\ell_0}, Z_{\ell_0}) = (x_{\ell_0}, z_{\ell_0})$) respectively. Moreover, we make the following additional assumption on the joint conditional probability density function $p_{X,L}$

Assumption 5.3.5.

(*i*) For all $x \in \mathbb{R}^d y \mapsto p_{X,L}(x,y)$ is continous on $\mathbb{R}$,

(*ii*) For all $\theta \in \mathbb{R}^d$, for every compact set $K \subset \mathbb{R}$ $\sup_{y\in K} p_{X,L}(x, \theta.x + y) \in L^1(\mathrm{d}x)$.

(*iii*) For all $\xi \in \mathbb{R}$, for every compact set $K \subset \mathbb{R}^d$ $\sup_{\theta\in K}(1+|x|^2)p_{X,L}(x,\theta.x+\xi) \in L^1(\mathrm{d}x)$,

(*iv*) $\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x > 0$, for all $(\xi^*_\alpha, \theta^*_\alpha) \in \operatorname{Arg\,min} V$.

Example 4. *We take the example (that will be studied in Section 5.5) of the energy provider which buys on an energy market a quantity* $C_T = \mu_C + \sigma_C G_1$ *where* $G_1 \sim \mathcal{N}(0,1)$ *of gas at price* S_T^g*, where gas spot price is modeled as a geometrice Brownian motion correlated with* $\rho \neq 0$ *to the consumption, namely*

$$S_T^g = S_0 e^{-\frac{\sigma_g^2}{2}T + \sigma_g\sqrt{T}\left(\rho G_1 + \sqrt{1-\rho^2}G_2\right)}$$

This quantity is sold to consumers at a price $K = S_0$*. The energy provider uses a one step self-financed strategy based on* S^g *to reduce its risk so that* $X = S_T^g - S_0$*. The loss* L *can be written*

$$L = (S_T^g - K)C_T = (S_T^g - S_0)C_T.$$

Using the change of variable formula, one shows that the joint conditional distribution function writes for $x > -S_0$*,* $x \neq 0$*,* $v \in \mathbb{R}$*,*

$$p_{X,L}(x,y) = \frac{1}{2\pi\rho\sigma_g\sigma_C\sqrt{T}}\frac{1}{(x+S_0)|x|}e^{-\frac{1}{2\rho^2}\left(\frac{1}{\sigma_g\sqrt{T}}\left(\log(\frac{x}{S_0}+1)+\frac{\sigma_g^2}{2}T\right)-\sqrt{1-\rho^2}\frac{1}{\sigma_C^2}\left(\frac{y}{x}-\mu_C\right)\right)^2}e^{-\frac{1}{2\sigma_C^2}\left(\frac{y}{x}-\mu_C\right)^2}.$$

Consequently, $p_{X,L}$ *satisfies* (i) *and* (iv)*. Moreover, for all* $x > -S_0$*,* $x \neq 0$ *and* y*,* $\theta \in \mathbb{R}$*,*

$$p_{X,L}(x,\theta.x+y) \leq \frac{1}{2\pi\rho\sigma_g\sigma_C\sqrt{T}}\frac{1}{(x+S_0)|x|} \in L^1(\mathrm{d}x).$$

so that, (ii) *is satisfied. Now, if* $K \subset \mathbb{R}$ *is a compact set, for all* $\xi \in \mathbb{R}$*, there exists a constant* $A > 0$ *such that for all* $x > -S_0$*,* $x \neq 0$

$$(1+|x|^2)p_{X,L}(x,\theta.x+\xi) \leq \frac{1}{2\pi\rho\sigma_g\sigma_C\sqrt{T}}\frac{1+|x|^2}{(x+S_0)|x|}e^{-A\frac{1}{x^2}} \in L^1(\mathrm{d}x),$$

so that (iii) *is satisfied.*

In next theorem, we use notations of 5.3.4. We establish a CLT for the empirical mean sequence $\bar{\phi}_n$ defined by (5.51).

Theorem 5.3.6. *(Convergence rate of the procedure) Suppose that Assumptions 5.2.4, 5.3.1 (with* $a > 1$*) and 5.3.5 are satisfied. If the step sequence is* $\gamma_n = \frac{\gamma_1}{n^\beta}$*, with* $\frac{1}{2} < \beta < 1$ *and* $\gamma_1 > 0$*, then the averaged procedure defined by* (5.51) *satisfies*

$$\sqrt{n}\left(\overline{\phi}_n - \phi^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma) \qquad as\ n \to +\infty$$

where the asymptotic covariance matrix Σ *is given by*

$$\Sigma = P^{-1}\Gamma(P^{-1})^T$$

with

$$P := \frac{1}{1-\alpha}\begin{pmatrix} \int_{\mathbb{R}^d} p_{X,L}(x,\xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x & \left(\int_{\mathbb{R}^d} x p_{X,L}(x,\xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^T & 0 \\ \int_{\mathbb{R}^d} x p_{X,L}(x,\xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x & \int_{\mathbb{R}^d} xx^T p_{X,L}(x,\xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x & 0 \\ 0 & 0 & 1-\alpha \end{pmatrix}, \tag{5.53}$$

and

$$\Gamma := \begin{pmatrix} \frac{\alpha}{1-\alpha} & 0 & \frac{\alpha\mathbb{E}\left[(L-\theta_\alpha^*.X-\xi_\alpha^*)_+\right]}{(1-\alpha)^2} \\ 0 & \frac{\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]}{(1-\alpha)^2} & \frac{\mathbb{E}\left[X(L-\theta_\alpha^*.X-\xi_\alpha^*)_+\right]}{(1-\alpha)^2} \\ \frac{\alpha\mathbb{E}\left[(L-\theta_\alpha^*.X-\xi^*)_+\right]}{(1-\alpha)^2} & \frac{\mathbb{E}\left[X(L-\theta_\alpha^*.X-\xi^*)_+\right]^T}{(1-\alpha)^2} & \frac{\mathrm{Var}\left((L-\theta_\alpha^*.X-\xi^*)_+\right)}{(1-\alpha)^2} \end{pmatrix}. \tag{5.54}$$

Proof. First note that the procedure (5.50) can be written

$$\forall\ n \geq 1, \quad \phi_n = \phi_{n-1} - \gamma_n \left(h(\phi_{n-1}) + \epsilon_n\right), \quad \phi_0 = (\xi_0, \theta_0, C_0),$$

where $h(\phi) := \mathbb{E}\left[H(\phi, U)\right] = \nabla V(\phi, x, z)$, and ϵ_n, $n \geq 1$, denotes the $\mathcal{F}_n$-adapted martingale increment sequence defined by for $i = 2, \cdots, d+1$:

$$\begin{aligned}
\epsilon_{1,n} &:= \frac{1}{1-\alpha}\left(\mathbb{P}\left(L - \theta.X \geq \xi\right)_{|\xi=\xi_{n-1},\theta=\theta_{n-1}} - \mathbf{1}_{\{L_n-\theta_{n-1}.X_n\geq\xi_{n-1}\}}\right), \\
\epsilon_{i,n} &:= \frac{1}{1-\alpha}\left(\mathbb{E}\left[X_{i-1}\mathbf{1}_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_{n-1},\theta=\theta_{n-1}} - X_{i-1,n}\mathbf{1}_{\{L_n-\theta_{n-1}.X_n\geq\xi_{n-1}\}}\right), \\
\epsilon_{d+2,n} &:= \Delta N_n = \frac{1}{1-\alpha}\left(\mathbb{E}\left[(L-\theta.X-\xi)_+\right]_{|\xi=\xi_{n-1},\theta=\theta_{n-1}} - (L_n-\theta_{n-1}.X_n-\xi_{n-1})_+\right),
\end{aligned}$$

where $L_n = F(U_n)$ and $X_n := G(U_n)$. Since the function V is convex, its hessian matrix P is positive as soon as h is differentiable. Now, in order to differentiate h, we write

$$\begin{aligned}
h_1(\phi) &= 1 - \frac{1}{1-\alpha}\int_{\mathbb{R}^d\times\mathbb{R}} p_{X,L}(x,y)\mathbf{1}_{\{y\geq\xi+\theta.x\}}\mathrm{d}x\mathrm{d}y, \\
h_i(\phi) &= -\frac{1}{1-\alpha}\int_{\mathbb{R}^d\times\mathbb{R}} x_i p_{X,L}(x,y)\mathbf{1}_{\{y\geq\xi+\theta.x\}}\mathrm{d}x\mathrm{d}y, \quad i = 2, \cdots, d+1, \\
h_{d+2}(\phi) &= C - \left(\xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\theta.X-\xi)_+\right]\right).
\end{aligned}$$

In order to differentiate h_1, note that, by Fubini's Theorem,

$$h_1(\phi) = 1 - \frac{1}{1-\alpha}\int_\xi^{+\infty}\mathrm{d}y\int_{\mathbb{R}^d} p_{X,L}(x,y)\mathrm{d}x = 1 - \frac{1}{1-\alpha}\int_{\mathbb{R}^d}\mathrm{d}x\int_{\xi+\theta.x}^{+\infty} p_{X,L}(x,y)\mathrm{d}y.$$

Owing to Assumption 5.3.5, one can interchange integral and derivation. In order to differentiate $h_{2:d+1}$, first note that, by Fubini's Theorem,

$$h_{2:d+1}(\phi) = -\frac{1}{1-\alpha}\int_\xi^{+\infty}\mathrm{d}y\int_{\mathbb{R}^d} x p_{X,L}(x,y)\mathrm{d}x = -\frac{1}{1-\alpha}\int_{\mathbb{R}^d}\mathrm{d}x\int_{\xi+\theta.x}^{+\infty} x p_{X,L}(x,y)\mathrm{d}y,$$

so that owing to Assumption 5.3.5 and Lebesgue's differentiation Theorem, one can interchange integral and derivation. Consequently, the functions h_1 and $h_{2:d+1}$ are

differentiable at $\phi^* := (\xi^*_\alpha, \theta^*_\alpha, C^*_\alpha)$ and for $i = 2, \cdots, d+1$,

$$\frac{\partial h_1}{\partial \xi}(\phi^*) = \frac{1}{1-\alpha}\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x, \frac{\partial h_1}{\partial C}(\phi^*) = \frac{\partial h_{d+2}}{\partial \xi}(\phi^*) = 0,$$

$$\begin{aligned}
\frac{\partial h_1}{\partial \theta_{i-1}}(\phi^*) = \frac{\partial h_i}{\partial \xi}(\phi^*) &= \frac{1}{1-\alpha}\int_{\mathbb{R}^d} x_i p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x, \\
\frac{\partial h_i}{\partial \theta_j}(\phi^*) = \frac{\partial h_j}{\partial \theta_i}(\phi^*) &= \frac{1}{1-\alpha}\int_{\mathbb{R}^d} x_i x_j p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x, \\
\frac{\partial h_i}{\partial C}(\phi^*) = \frac{\partial h_{d+2}}{\partial \theta_i}(\phi^*) &= 0, \quad \frac{\partial h_{d+2}}{\partial C}(\phi^*) = 1,
\end{aligned}$$

so that M is given by (5.53). Let $u = (u_1, u_2, u3) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$,

$$\begin{aligned}
u^T M u = \frac{\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x}{1-\alpha} \Bigg(u_1^2 + 2u_1 \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x}\mathrm{d}x u_2 \\
+ u_2^T \frac{\int_{\mathbb{R}^d} x x^T p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x} u_2 + u_3^2 \Bigg),
\end{aligned}$$

using the inequality $2u_1 \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x}\mathrm{d}x u_2 \geq -u_1^2 - u_2^T \mathbb{E}_{\mathbb{Q}}[X]\, \mathbb{E}_{\mathbb{Q}}[X]^T u_2$, we obtain

$$\begin{aligned}
u^T M u \geq \frac{\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x}{1-\alpha} \Bigg(u_2^T \int_{\mathbb{R}^d} \Bigg(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x} \Bigg) \\
\times \Bigg(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x} \Bigg)^T p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)\mathrm{d}x u_2 + u_3^2 \Bigg), \\
> 0.
\end{aligned}$$

Consequently, the matrix P is a uniformly repulsive matrix. To apply Theorem 5.3.4, we need to check assumptions (RP) of (5.52). Let $C > 0$. First note that

$$\mathbb{E}\left[|\epsilon_{1,n+1}|^{2a} | \mathcal{F}_n\right] \mathbf{1}_{\{|\phi_n - \phi^*| \leq C\}} \leq \left(\frac{1}{1-\alpha}\right)^{2a} 2^{2a} < +\infty.$$

Thanks to Assumtpion 5.2.4 (with $a > 1$), there exists $A > 0$, such that for $i = 2, \cdots, d+1$,

$$\mathbb{E}\left[|\epsilon_{i,n+1}|^{2a} | \mathcal{F}_n\right] \mathbf{1}_{\{|\phi_n - \phi^*| \leq C\}} \leq A\mathbb{E}\left[X_{i-1}^{2a}\right] < +\infty,$$

and

$$\mathbb{E}\left[|\epsilon_{d+2,n+1}|^{2a} | \mathcal{F}_n\right] \mathbf{1}_{\{|\phi_n - \phi^*| \leq C\}} \leq A\left(\mathbb{E}\left[|L|^{2a}\right] + \mathbb{E}\left[|X|^{2a}\right]\right) < +\infty.$$

Consequently, (ii) of (5.52) holds true with $b = 2a$ since

$$\sup_{n \geq 0} \mathbb{E}\left[|\epsilon_{n+1}|^{2a} | \mathcal{F}_n\right] \mathbf{1}_{\{|\phi_n - \phi^*| \leq C\}} < +\infty.$$

It remains to check (iii) for some positive definite symmetric matrix Γ.

The continuity of the functions $(\xi,\theta)\mapsto \mathbb{E}\left[X_{i-1}X_{j-1}\mathbf{1}_{\{L-\theta.X\geq\xi\}}\right]$ and $(\xi,\theta)\mapsto \mathbb{E}\left[X_{i-1}\mathbf{1}_{\{L-\theta.X\geq\xi\}}\right]$ at $(\xi^*_\alpha,\theta^*_\alpha)$ which follows from the continuity of the joint distribution (L,X), combined with the equality $\mathbb{E}\left[X_{i-1}\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right]=0$, $i=2,\cdots,d+1$, implies that

$$\begin{aligned}\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{i,j}|\mathcal{F}_n\right] &= \mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{j,i}|\mathcal{F}_n\right]\\ &= \frac{1}{(1-\alpha)^2}\Big(\mathbb{E}\left[X_{i-1}X_{j-1}\mathbf{1}_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_n,\theta=\theta_n}-\\ &\quad \mathbb{E}\left[X_{i-1}\mathbf{1}_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_n,\theta=\theta_n}\mathbb{E}\left[X_{j-1}\mathbf{1}_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_n,\theta=\theta_n}\Big),\\ &\xrightarrow{a.s.} \frac{1}{(1-\alpha)^2}\mathbb{E}\left[X_{i-1}X_{j-1}\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right].\end{aligned}$$

Using similar arguments one shows that $\mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right]\xrightarrow{a.s.}\Gamma$. This completes the proof. □

One may be interested by the asymptotic variance of each components of the algorithm, namely ξ_n, θ_n and C_n rather than the whole asymptotic matrix. The inverse matrix P^{-1} can be written

$$P^{-1}:=\frac{1-\alpha}{\int_{\mathbb{R}^d}p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}\begin{pmatrix}1+V^T\Pi^{-1}V & -V^T\Pi^{-1} & 0\\ -\Pi^{-1}V & \Pi^{-1} & 0\\ 0 & 0 & \frac{1}{1-\alpha}\int_{\mathbb{R}^d}p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x\end{pmatrix}, \tag{5.55}$$

where $\Pi:=\int_{\mathbb{R}^d}\left(x-\frac{\int_{\mathbb{R}^d}xp_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)}{\int_{\mathbb{R}^d}p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}\right)\left(x-\frac{\int_{\mathbb{R}^d}xp_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)}{\int_{\mathbb{R}^d}p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}\right)^T p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x$, and $V:=\frac{1}{\int_{\mathbb{R}^d}p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}\int_{\mathbb{R}^d}xp_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x$, so that, for $i=2,\cdots,d+1$,

$$\begin{aligned}\Sigma_{1,1} &= \frac{1}{\left(\int_{\mathbb{R}^d}p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x\right)^2}\Big(\left(1+V^T\Pi^{-1}V\right)^2\alpha(1-\alpha)\\ &\quad+\left(\Pi^{-1}V\right)^T\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right]\Pi^{-1}V\Big),\end{aligned} \tag{5.56}$$

$$\Sigma_{i,i} = \frac{1}{\left(\int_{\mathbb{R}^d}p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x\right)^2}\Big(m_i^2\alpha\left(1-\alpha\right)+\tilde{m}_{.i}^T\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right]\tilde{m}_{.i}\Big), \tag{5.57}$$

$$\Sigma_{d+2,d+2} = \left(\frac{1}{1-\alpha}\right)^2\mathrm{Var}\left(\left(L-\theta^*_\alpha.X-\xi^*_\alpha\right)_+\right), \tag{5.58}$$

where $m=\Pi^{-1}V$ and $\Pi^{-1}=(\tilde{m}_{i,j})_{1\leq i\leq d,1\leq j\leq d}$.

5.3.3 Dynamic CVaR hedging

In this section, we propose several methods to compute the optimal strategy of (5.1), the VaR and the CVaR of the CVaR-hedged portfolio. From a modeling point of view, under Assumption 5.1.2, we suppose for every $\ell=1,\cdots,M$ that there exists

two continuous functions $G_\ell : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_\ell} \to \mathbb{R}^d$, $F_\ell : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_\ell} \to \mathbb{R}$ such that

$$X_\ell - X_{\ell-1} = G_\ell(X_{\ell-1}, Z_{\ell-1}, U_\ell), \quad \text{and} \quad L = F_\ell\left(X_{\ell-1}, Z_{\ell-1}, U_\ell\right),$$

where U_ℓ is a r_ℓ-dimensional random variable independent of $\mathcal{G}_{l-1}$.

Crude CVaR hedging Algorithm (C.H.)
The direct approach to solve (5.1) is to proceed as in the static framework and to devise a global stochastic gradient algorithm. To be more precise, at every time t_k, we consider an optimal N_k-quantization $\left(\hat{X}_k, \hat{Z}_k\right)$, $k = 1, \cdots, M-1$, based on an optimal quantization N_k grid $\Gamma_k = \left(\left(x_k^1, z_k^1\right), \cdots, \left(x_k^{N_k}, z_k^{N_k}\right)\right)$ of the state process at time t_k.

A careful reading of Section 5.2.4 shows that the optimal number of shares to be held over the time period $(k, k+1]$, θ_k^* depends of the whole process (X, Z). For this method, we make the approximation which consists of making θ_k^* depending only of the state process at time k, (X_k, Z_k). Thus, we only need to estimate $\theta_k^{j,*}$ at each time t_k, for all nodes $\left(x_k^j, z_k^j\right)$, $j = 1, \cdots, N_k$ on the corresponding grid. This can be done by the following stochastic algorithm, namely,

$$\xi_n = \xi_{n-1} - \gamma_n H_1\left(\xi_{n-1}, \theta_{n-1}, U_n\right), \tag{5.59}$$
$$\theta_{0,n} = \theta_{0,n-1} - \gamma_n H_{2,0}\left(\xi_{n-1}, \theta_{n-1}, U_n\right), \tag{5.60}$$
$$\theta_{\ell,n}^j = \theta_{\ell,n-1}^j - \gamma_n H_{2,\ell}^j\left(\xi_{n-1}, \theta_{n-1}, U_n\right), \ j = 1, \cdots, N_\ell, \ k = 1, \cdots, M-1, \tag{5.61}$$
$$C_n = C_{n-1} - \gamma_n H_3\left(\xi_{n-1}, \theta_{n-1}, C_{n-1}, U_n\right), \tag{5.62}$$

where $\theta_n = (\theta_{0,n}, \cdots, \theta_{M-1,n})$, $(U_n)_{n\geq 1} = \left((U_{1,n}, \cdots, U_{M,n})_{n\geq 1}\right)$ are i.i.d. random variables with $U_{\ell,n} \sim U_\ell$ and for $\ell = 1, \cdots, M$ and $j = 1, \cdots, N_\ell$ the functions H_1, H_3 and $H_{2,\ell}^j$ are defined by

$$H_1(\xi, \theta, u) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\left\{\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i \geq \xi\right\}},$$
$$H_{2,0}\left(\xi, \theta, u\right) = -\frac{G_1(X_0, Z_0, U_1)}{1-\alpha}\mathbf{1}_{\left\{\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i \geq \xi\right\}},$$
$$H_{2,\ell}^j\left(\xi, \theta, u\right) = -\frac{G_{\ell+1}(X_\ell, Z_\ell, U_{\ell+1})}{1-\alpha}\mathbf{1}_{\left\{\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i \geq \xi\right\}}\mathbf{1}_{\{(X_\ell, Z_\ell) \in C_j(x_\ell, z_\ell)\}},$$
$$H_3\left(\xi, \theta, C, u\right) = C - \xi - \frac{1}{1-\alpha}\left(\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i - \xi\right)_+.$$

The sequence $(\xi_n, \theta_n, C_n)_{n\geq 1}$ *a.s.* converges toward its target $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$. Note that the dimension of the sequence (ξ_n, θ_n, C_n) to be updated at each step of the algorithm is equal to $D := 2 + d + \sum_{\ell=2}^M d \times N_\ell$.

When the dimension is low ($D \leq 100$), which is often due to the fact that the number of trading dates is low (say $M \leq 5$) and the number of traded assets used for the hedging is small ($d \approx 1, 2$), the above algorithm is very efficient and we observe a great reduction of the CVaR compared to the static case (5.3).

However, if we consider a portfolio with a time horizon $T = 1$ year, 12 trading dates (one each month), if the investor hedges using 5 stocks and in the case where all

layers in the quantization grids have the same size, *i.e.* $N = N_\ell = 5$, $\ell = 1, \cdots, M$, the dimension of the algorithm is $D = 282$. This example is a reasonable case in the energy sector when an energy company has to provide electricity or gas to consumers all year long and simultaneously needs to control and hedge its risk every month using electricity and/or gas forward contracts since the underlying spot is not storable. For instance, one may use 12 forward contracts with maturity $T_\ell = t_\ell$ that delivers electricity or gas over a period which corresponds to each month of the considered year. From a practical point of view, Electricity and Gas Futures market enable to trade: the next three months, the next two quarters and the next three electricity or gas seasons (see the Powernext Gas Futures market for instance), thus one may proceed to a rough risk hedging using only some of these contracts. However, when dealing with a portfolio that depends on several energy commodities as it is often the case in energy market, the dimension of the considered RM algorithm becomes a real issue.

From a numerical point of vue, we observe that in a high dimensional framework the algorithm "freezes" and "suffers", say as soon as the dimension is greater than 100 or 150. Moreover, we observe that some components of θ_n are never updated by the algorithm. That is the bottleneck of this first algorithm in pratical implementation. To overcome this problem, we propose several approximate solutions to solve (5.1) which crucially relies on Assumptions 5.1.1, 5.1.2 and 5.2.8. These solutions have the major advantage to dramatically reduce the dimension of the above algorithm.

Backward dynamic hedging strategy (B.H.)
This strategy is based on (5.9) and consists in a backward resolution. To be more precise, if we consider M trading dates, then (5.9) and Assumption 5.1.1 imply that in order to hedge the risk at the last trading date t_{M-1}, we have to solve

$$\inf_{\theta\in\mathcal{A}} \mathbb{E}\left[\mathcal{G}_{M-1}\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right] = \inf_{\theta_{M-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{t_{M-1}},\mathbb{P})} \mathbb{E}\left[\mathcal{G}_{M-1}\text{-CVaR}_\alpha\left(L - \theta_{M-1}.\Delta X_M\right)\right].$$

The optimization problem that appears in the right-hand side of the above equalitie can be easily solved using the static algorithm developed in Section 3. Now that we have the solution θ^b_{M-1} of this problem, we can go one step backward and solve the new problem

$$\inf_{\theta_{M-2}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{t_{M-2}},\mathbb{P})} \mathbb{E}\left[\mathcal{G}_{M-2}\text{-CVaR}_\alpha\left(L - \theta^b_{M-1}.\Delta X_M - \theta_{M-2}\Delta X_{M-1}\right)\right],$$

using again the algorithm developed in the static framework in order to obtain θ^b_{M-2}. Following this idea till time 0, we obtain step by step the backward hedging strategy $\theta^b \equiv \left(\theta^b_\ell\right)_{1\le\ell\le M-1}$.

Although this method is not optimal from a theoretical point of view, it has the advantage to provide a strategy which controls the risk at each time step until maturity. However we observe on numerical experiments that the resulting static CVaR related to this self-financed strategy θ^b, namely

$$\text{CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta^b_{\ell-1}.\Delta X_\ell\right),$$

is significantly higher than the one obtained by the first global algorithm (C.H.). The reason is that by solving at step $k+1$ the optimization problem

$$\inf_{\theta_{M-k-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{M-k-1},\mathbb{P})} \mathbb{E}\left[\mathcal{G}_{M-k-1}\text{-CVaR}_\alpha\left(L-\sum_{\ell=M-k+1}^{M}\theta^b_{\ell-1}.\Delta X_\ell-\theta_{M-k-1}\Delta X_{M-k-1}\right)\right],$$

there is an error (compared to the original problem (5.1)) on the estimate $\theta^b_{M-\ell-1}\neq\theta^*_{M-\ell-1}$ which propagates at each step and can become more and more important as the number of trading dates increases. That is the major drawback of this procedure.

Dynamic hedging strategy based on a martingale decomposition of L (M.D.H.)
This method is based on the sub-additivity of the CVaR and on the following decomposition of the loss L into a sum of $\mathbb{G}$-martingale increments, namely

$$L=\mathbb{E}[L]+\sum_{\ell=1}^{M}\tilde{\Delta}L_\ell, \tag{5.63}$$

where $\tilde{\Delta}L_\ell=\mathbb{E}[L|\,\mathcal{G}_\ell]-\mathbb{E}[L|\,\mathcal{G}_{\ell-1}]$, $1\leq\ell\leq M$. Now, using the sub-additivity of the CVaR, we obtain

$$\inf_{\theta\in\mathcal{A}}\text{CVaR}_\alpha\left(L-\sum_{\ell=1}^{M}\theta_{\ell-1}\Delta X_\ell\right)\leq\mathbb{E}[L]+\sum_{\ell=1}^{M}\inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1},\mathbb{P})}\text{CVaR}_\alpha\left(\tilde{\Delta}L_\ell-\theta_{\ell-1}\Delta X_\ell\right). \tag{5.64}$$

The right-hand side of the last inequality shows that for each time step we have to solve a one step static local CVaR-hedging problem. From a numerical point of view, indeed, when the dimension of the algorithm is not too large (say $D\leq 150$), we observe that the CVaR obtained using this strategy is almost equal to the optimal one. When the dimension D becomes large, which is generally due to a large number of trading dates, we observe a good behavior with a real improvement on the CVaR when the number of trading dates increases. An even better behavior is obtained by slightly modifying this second aproach. We use the inequality

$$\inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1},\mathbb{P})}\mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-CVaR}_\alpha\left(\tilde{\Delta}L_\ell-\theta_{\ell-1}\Delta X_\ell\right)\right]\leq\inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1},\mathbb{P})}\text{CVaR}_\alpha\left(\tilde{\Delta}L_\ell-\theta_{\ell-1}\Delta X_\ell\right), \tag{5.65}$$

and switch to the new optimization problem

$$\sum_{\ell=1}^{M}\inf_{\theta_{\ell-1}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell-1},\mathbb{P})}\mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-CVaR}_\alpha\left(\tilde{\Delta}L_\ell-\theta_{\ell-1}\Delta X_\ell\right)\right]. \tag{5.66}$$

The main difference in solving the local problem in the left hand-side of (5.65) compared to the right hand-side appears in the variable ξ_n of the two associated RM algorithms, which corresponds to the estimate at step n of the VaR_α. In this new version, like in the static case, the variable ξ_n is local and depends of the considered nodes whereas in the other version this variable is *global* and is the same for all nodes of the quantization tree.

Although, to our knowledge, there is neither an equality nor an inequality between the original problem and (5.66), numerical experiments led us to the conclusion that this last algorithm behaves better than the one obtained by solving the right-hand side of (5.65) at each time step. To be more precise, the original CVaR estimated by the strategy obtained by (5.66) is lower than the one obtained by the strategy solution of the right-hand side in (5.64).

In order to solve (5.66), we use optimal vector quantization again to approximate the unknown random variable $\tilde{\Delta}L_\ell$, *i.e.*, we approximate $\mathbb{E}\left[L|\mathcal{G}_\ell\right]$ by using the cubature formula

$$\mathbb{E}\left[F_{\ell+1}(X_\ell, Z_\ell, U_{\ell+1})|\,(X_\ell, Z_\ell)\right] \approx \varphi(X_\ell, Z_\ell) = \sum_{j=1}^{N_\ell} F(X_\ell, Z_\ell, u_{\ell+1}^j)\mathbb{P}\left(U_{\ell+1} \in C_j(u_{\ell+1})\right), \tag{5.67}$$

and design at each time step a RM algorithm based on the procedure investigated in the static framework. One may have considered the classical decomposition

$$L = L_0 + \sum_{\ell=1}^{M} \Delta L_\ell \tag{5.68}$$

instead of (5.63). However, it is quite natural to approximate from the sequence of martingale increments ΔX_ℓ, $\ell = 1, \cdots, M$, another martingale sequence so that the decomposition (5.63) is more appropriate to our framework than the decomposition (5.68). The method based on this classical decomposition of L will be called *C.D.H.*

5.4 Design of faster procedures: variance reduction techniques

In practice, the convergence of the different considered algorithms (static and dynamic framework) will be slow and chaotic when the confidence level α is close to 1. This is due to the fact they are only updated on rare events since it tries to measure the tail distribution: $\mathbb{P}\left(L - \theta_\alpha^*.X > \xi_\alpha^*\right) = 1 - \alpha \approx 0$. Another problem may be the simulation of L and X. Each evaluation may require a lot of computational efforts and takes a long time. So, for practical implementation it is necessary to combine the above procedures with variance reduction techniques to achieve accurate estimates at a reasonable cost.

In this section, we develop two variance reduction techniques in order to reduce the asymptotic variance in the CLT (5.53). The first one is based on the unconstrained importance sampling (IS) stochastic algorithm developed in [62] and then applied to both VaR and CVaR in [6]. We show how it can be combined adaptively with our algorithms. Consequently, every new sample is used to dynamically optimize the IS change of measure and the estimate of $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$. This kind of algorithm is known to be a powerfull tool when dealing with rare events. The second one is based on LCV. We use a control variable based on X, since Assumption 5.1.1 implies $\mathbb{E}\left[X\right] = 0$. For sake of simplicity, we only develop those two methods in the static self-financed strategy framework though it can be easily generalized to the other considered algorithm.

5.4.1 Unconstrained Recursive Importance Sampling

In this section, we study the classical IS by translation. Since the procedure is very similar to the one applied in [6], we only give results without any proof.

Applied to the Monte-Carlo computation of $\mathbb{E}[F(Y)]$, where Y is an $\mathbb{R}^d$-valued random variable, $F \in L^2(\mathbb{P}_Y)$ satisfies $\mathbb{P}(F(Y) \neq 0) > 0$ and Y has a probability density function p, it consists in using the invariance of the Lebesgue measure by translation, for every $\mu \in \mathbb{R}^d$,

$$\mathbb{E}[F(Y)] = \mathbb{E}\left[F(Y+\mu)\frac{p(Y+\mu)}{p(Y)}\right], \tag{5.69}$$

and among all these random variables with the same expectation in selecting the one with the lowest variance, *i.e.* the one with the lowest quadratic norm

$$\min_{\mu\in\mathbb{R}^d}\left\{Q(\mu) := \mathbb{E}\left[F^2(Y+\mu)\frac{p^2(Y+\mu)}{p^2(Y)}\right] = \mathbb{E}\left[F^2(Y)\frac{p(Y)}{p(Y-\mu)}\right]\right\}.$$

Applied to our framework, we plan to minimize the asymptotic variance of each components of the algorithm (in its averaged form, as detailed above) *i.e.* each $\Sigma_{i,i}$, $i = 1, \cdots, d+2$ which appears in the CLT (5.53). However, the two matrix Π^{-1} and $\Pi^{-1}V$ are intrinsic constant (and comes in fact from the Jacobian matrix $Dh(\xi^*_\alpha, \theta^*_\alpha, C^*_\alpha)$ of the mean function h of the algorithm and cannot be change by applying IS unless we deeply modify the nature of the original algorithm $(\xi_n, \theta_n, C_n)_{n\geq 1}$.

The asymptotic variance of the VaR procedure $\Sigma_{1,1}$ in (5.56) presents two terms:

$$\left(1 + V^T\Pi^{-1}V\right)^2 \alpha(1-\alpha) = \left(1 + V^T\Pi^{-1}V\right)^2 \mathrm{Var}\left(\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right)$$

corresponds to the "intrinsic" asymptotic variance of the VaR procedure, and

$$\Pi^{-1}V^T\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right]\Pi^{-1}V,$$

corresponds to the influence of the regression procedure. The asymptotic variance obtained in the above equation cannot be minimized directly by IS since it is a quadratic form in the unknown vector $\Pi^{-1}V^T$. In order to devise an IS algorithm, we write

$$\Pi^{-1}V^T\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right]\Pi^{-1}V \leq \left|\Pi^{-1}V\right|^2\mathbb{E}\left[|X|^2\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right],$$

The asymptotic variances $\Sigma_{i,i}$, $i = 2, \cdots, d+1$, appear quite similar so that we can obtain similar inequalities. Thus, in order to improve the convergence of the VaR_α and the θ^*_α components we aim at minimizing by IS the two quantities

$$\mathrm{Var}\Big(\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\Big), \quad \text{and} \quad \mathbb{E}\left[|X|^2\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right] = \sum_{i=1}^{d}\mathbb{E}\left[X_i^2\mathbf{1}_{\{L-\theta^*_\alpha.X\geq\xi^*_\alpha\}}\right]$$

provided the non-degeneracy assumption

$$\forall(\xi,\theta) \in \mathrm{Arg\,min}\, V, \quad \mathbb{P}\left(X_i\mathbf{1}_{\{L-\theta.X\geq\xi\}} \neq 0\right) > 0, \quad i = 1, \cdots, d, \tag{5.70}$$

holds. From a theoretical point of view, it is not optimal but our numerical experiments indicate that minimizing those terms provides less variance during the first

iterations than the original procedure without variance reduction. As concerns the CVaR_α, we can minimize directly the asymptotic variance

$$\mathrm{Var}\left((L-\theta_\alpha^*.X-\xi_\alpha^*)_+\right).$$

Now, we assume that U has an absolutely continuous distribution $\mathbb{P}_U(du)=p(u)\lambda_r(du)$ where λ_r denotes the Lebesgue measure on $(\mathbb{R}^r,\mathcal{B}or(\mathbb{R}^r))$.

To minimize the three variances $\mathrm{Var}\Big(\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\Big)$, $\mathbb{E}\left[X_i^2\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]$ and $\mathrm{Var}\left((L-\theta_\alpha^*.X-\xi_\alpha^*)_+\right)$ respectively, we use the invariance of the Lebesgue measure by translation for $\mu_i\in\mathbb{R}^r$, $i=1,\cdots,d+2$ as follows,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]&=\mathbb{E}\left[\mathbf{1}_{\left\{L^{(+\mu_1)}-\theta_\alpha^*.X^{(+\mu_1)}\geq\xi_\alpha^*\right\}}\frac{p(U+\mu_1)}{p(U)}\right],\\
\mathbb{E}\left[X_i\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]&=\mathbb{E}\left[X_i^{(+\mu_i)}\mathbf{1}_{\left\{L^{(+\mu_i)}-\theta_\alpha^*.X^{(+\mu_i)}\geq\xi_\alpha^*\right\}}\frac{p(U+\mu_i)}{p(U)}\right],\ i=2,\cdots,d+1,\\
\mathbb{E}\left[(L-\theta_\alpha^*.X-\xi_\alpha^*)_+\right]&=\mathbb{E}\left[\left(L^{(+\mu_{d+2})}-\theta_\alpha^*.X^{(+\mu_{d+2})}-\xi_\alpha^*\right)_+\frac{p(U+\mu_{d+2})}{p(U)}\right],
\end{aligned}$$

where for sake of simplicity $L^{(\pm\mu)}=F(U\pm\mu)=F(x,z,U\pm\mu)$, $X^{(\pm\mu)}=G(U\pm\mu)=G(x,z,U\pm\mu)-x$ and $X_i^{(\pm\mu)}=G_i(U\pm\mu)-x_i$, for $\mu_i\in\mathbb{R}^r$. Let us temporarily forget that of course we do not know $(\xi_\alpha^*,\theta_\alpha^*)$ at this stage. Among all these random random variables with the same expectation, we want to select the one with the lowest quadratic norms for $\mu=(\mu_1,\cdots,\mu_{d+2})\in(\mathbb{R}^r)^{d+2}$,

$$\begin{aligned}
Q_1(\mu_1,\xi_\alpha^*,\theta_\alpha^*)&:=\mathbb{E}\left[\mathbf{1}_{\left\{L^{(+\mu_1)}-\theta_\alpha^*.X^{(+\mu_1)}\geq\xi_\alpha^*\right\}}\frac{p^2(U+\mu_1)}{p^2(U)}\right],\\
Q_i(\mu_i,\xi_\alpha^*,\theta_\alpha^*)&:=\mathbb{E}\left[\left(X_i^{(+\mu_i)}\right)^2\mathbf{1}_{\left\{L^{(+\mu_i)}-\theta_\alpha^*.X^{(+\mu_i)}\geq\xi_\alpha^*\right\}}\frac{p^2(U+\mu_i)}{p^2(U)}\right],\\
Q_{d+2}(\mu_{d+2},\xi_\alpha^*,\theta_\alpha^*)&:=\mathbb{E}\left[\left(L^{(+\mu_{d+2})}-\theta_\alpha^*.X^{(+\mu_{d+2})}-\xi_\alpha^*\right)_+^2\frac{p^2(U+\mu_{d+2})}{p^2(U)}\right],
\end{aligned}$$

If the following assumption

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{L-\theta.X\geq\xi\}}\frac{p(U)}{p(U-\mu)}\right]&<+\infty,\\
\mathbb{E}\left[X_i^2\mathbf{1}_{\{L-\theta.X\geq\xi\}}\frac{p(U)}{p(U-\mu)}\right]&<+\infty,\\
\mathbb{E}\left[(L-\theta.X-\xi)_+^2\frac{p(U)}{p(U-\mu)}\right]&<+\infty,
\end{aligned}\tag{5.71}$$

holds true, for all $\mu\in\mathbb{R}^r$, for all $(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^r$, then, for $i=1,\cdots,d+2$, Q_i is everywhere finite and a reverse change of variable shows that for every $\mu\in\mathbb{R}^r$,

$$Q_1(\mu,\xi_\alpha^*,\theta_\alpha^*)=\mathbb{E}\left[\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\frac{p(U)}{p(U-\mu)}\right],\tag{5.72}$$

$$Q_i(\mu,\xi_\alpha^*,\theta_\alpha^*)=\mathbb{E}\left[X_i^2\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\frac{p(U)}{p(U-\mu)}\right],\tag{5.73}$$

$$Q_{d+2}(\mu,\xi_\alpha^*,\theta_\alpha^*)=\mathbb{E}\left[(L-\theta_\alpha^*.X-\xi_\alpha^*)_+^2\frac{p(U)}{p(U-\mu)}\right].\tag{5.74}$$

Now if p satisfies

$$\begin{cases} (i) \quad \forall u \in \mathbb{R}^r,\ \mu \mapsto p(u-\mu) \text{ is log-}concave, \\ (ii) \quad \forall u \in \mathbb{R}^r,\ \lim_{|\mu|\to+\infty} p(u-\mu) = 0 \quad \text{or } \forall u \in \mathbb{R}^r,\ \lim_{|\mu|\to+\infty} \frac{p(u-\mu)}{p^2(u-\frac{\mu}{2})} = 0, \end{cases} \tag{5.75}$$

one shows that, for $i = 1, \cdots, d+2$, $Q_i(., \xi^*_\alpha, \theta^*_\alpha)$ is (strictly) finite, convex, goes to infinity at infinity so that $\arg\min Q_i(., \xi^*_\alpha, \theta^*_\alpha) = \{\mu_i \in \mathbb{R}^r \mid \nabla_{\mu_i} Q_i(\mu_i, \xi^*_\alpha, \theta^*_\alpha) = 0\}$ is non empty (see [62]). In order to devise a recursive RM procedure to approximate the optimal parameters $\mu^*_{i,\alpha}$, $i = 1, \cdots, d+2$, we are led to differentiate $Q_i(., \xi^*_\alpha, \theta^*_\alpha)$. We introduce the following assumption on the probability density p of U

$$\exists b \in [1,2] \text{ such that } \begin{cases} (i) \quad \frac{|\nabla p(u)|}{p(u)} = O(|u|^{b-1}) \quad \text{as} \quad |u| \to \infty \\ (ii) \quad \exists \rho > 0, u \mapsto \log(p(u)) + \rho|u|^b \text{ is convex,} \end{cases} \tag{5.76}$$

and introduce the assumption on $F(U)$ and $G(U)$

$$\forall C > 0, \quad \mathbb{E}\left[\left(1 + F(U)^2 + |G(U)|^2\right) e^{C|U|^{b-1}}\right] < +\infty. \tag{5.77}$$

One shows that as soon as (5.70), (5.71), (5.75), (5.76), (5.77) are satisfied, Q_i is finite and differentiable on $\mathbb{R}^r$ with a gradient given by

$$\nabla Q_1(\mu, \xi^*_\alpha, \theta^*_\alpha) = \mathbb{E}\left[\mathbf{1}_{\left\{L^{(-\mu)} - \theta^*_\alpha . X^{(-\mu)} \geq \xi^*_\alpha\right\}} \underbrace{\frac{p^2(U-\mu)}{p(U)p(U-2\mu)} \frac{\nabla p(U-2\mu)}{p(U-2\mu)}}_{:=W(\mu,U)}\right], \tag{5.78}$$

$$\nabla Q_i(\mu, \xi^*_\alpha, \theta^*_\alpha) = \mathbb{E}\left[\left(X_i^{(-\mu)}\right)^2 \mathbf{1}_{\left\{L^{(-\mu)} - \theta^*_\alpha . X^{(-\mu)} \geq \xi^*_\alpha\right\}} W(\mu, U)\right], \tag{5.79}$$

$$\nabla Q_{d+2}(\mu, \xi^*_\alpha, \theta^*_\alpha) = \mathbb{E}\left[\left(L^{(-\mu)} - \theta^*_\alpha . X^{(-\mu)} - \xi^*_\alpha\right)_+^2 W(\mu, U)\right], \tag{5.80}$$

for all $\mu \in \mathbb{R}^r$. This expression may look complicated at first glance but in fact the weight term $W(\mu, U)$ can be easily controlled by a deterministic function of μ since

$$|W(\mu, u)| \leq e^{2\rho|\mu|^b}(A|u|^{b-1} + A|\mu|^{b-1} + B), \tag{5.81}$$

for some real constants A and B. In the case of a normal distribution $U \overset{d}{=} \mathcal{N}(0;1)$,

$$W(\mu, U) = e^{\mu^2}(2\mu - U).$$

Now if we have a control on the growth of the function F and G, typically for some positive constants C and c

$$\begin{cases} \forall u \in \mathbb{R}^r, |F(u)| \leq \tilde{F}(u) \quad \text{and} \quad \tilde{F}(u+v) \leq C(1 + \tilde{F}(u))^c (1 + \tilde{F}(v))^c, \\ \forall u \in \mathbb{R}^r, |G(u)| \leq \tilde{G}(u) \quad \text{and} \quad \tilde{G}(u+v) \leq C(1 + \tilde{G}(u))^c (1 + \tilde{G}(v))^c, \\ \mathbb{E}\left[|U|^{2(b-1)}\left(\tilde{F}(U)^{4c} + \tilde{G}(U)^{4c}\right) + \tilde{G}(U)^{4c}\right] < +\infty, \end{cases} \tag{5.82}$$

then we can define, for $\mu \in (\mathbb{R}^r)^{d+2}$ and $i = 2, \cdots, d+1$

$$K_1(\mu_1, \xi_\alpha^*, \theta_\alpha^*, U) = e^{-2\rho|\mu_1|^b} \mathbf{1}_{\left\{L^{(-\mu_1)} - \theta_\alpha^* . X^{(-\mu_1)} \geq \xi_\alpha^*\right\}} W(\mu_1, U), \tag{5.83}$$

$$K_i(\mu_i, \xi_\alpha^*, \theta_\alpha^*, U) = \frac{e^{-2\rho|\mu_i|^b}}{1 + \tilde{G}(-\mu_i)^{2c}} \left(X_i^{(-\mu_i)}\right)^2 \mathbf{1}_{\left\{L^{(-\mu_i)} - \theta_\alpha^* . X^{(-\mu_i)} \geq \xi_\alpha^*\right\}} W(\mu_i, U), \tag{5.84}$$

$$K_{d+2}(\mu_{d+2}, \xi_\alpha^*, \theta_\alpha^*, U) = \frac{e^{-2\rho|\mu_{d+2}|^b}}{1 + \tilde{F}(-\mu_{d+2})^{2c} + |\theta_\alpha^*|^{2c} \tilde{G}(-\mu_{d+2})^{2c}} \times \left(L^{(-\mu_{d+2})} - \theta_\alpha^* . X^{(-\mu_{d+2})} - \xi_\alpha^*\right)_+^2 W(\mu_{d+2}, U), \tag{5.85}$$

so that it satisfies the linear growth assumption (2.8) of the RM Theorem and

$$\{\mu_i \in \mathbb{R}^r \mid \mathbb{E}\left[K_i(\mu_i, \xi_\alpha^*, \theta_\alpha^*, U)\right] = 0\} = \{\mu_i \in \mathbb{R}^r \mid \nabla_{\mu_i} Q_i(\mu_i, \xi_\alpha^*, \theta_\alpha^*) = 0\}.$$

Moreover, since Q_i is convex $\nabla_{\mu_i} Q_i$ satisfies (2.7). Now the RM algorithms defined for $n \geq 1$ by

$$\mu_{i,n} = \mu_{i,n-1} - \gamma_n K_i(\mu_{i,n-1}, \xi_\alpha^*, \theta_\alpha^*, U_n), \ \mu_{i,0} \in \mathbb{R}^r,$$

a.s. converges to an $\operatorname{Arg\,min} Q_i(., \xi_\alpha^*, \theta_\alpha^*)$ (square integrable) random variable $\mu_{i,\alpha}^*$ (for more details about unconstrained recursive IS, we refer to [62] and [6]). Now, since we do not know either ξ_α^* and θ_α^* respectively, we make the whole procedure adaptive by replacing at step n, these unknown parameters by their running approximation at step $n-1$. This finally justifies to introduce the following global procedure. One defines the state variable, for $n \geq 0$,

$$\phi_n = (\xi_n, \theta_n, C_n, \mu_{1,n}, \cdots, \mu_{d+2,n}), \tag{5.86}$$

where ξ_n, θ_n, C_n denotes the VaR_α, the regression vector and the CVaR_α estimates at step n, μ_1 denotes the variance reducer for the VaR_α, μ_i denotes the variance reducer for the ith component of θ_α^*, *i.e.* $\theta_{i,\alpha}^*$ and μ_{d+2} denotes the variance reducer for the CVaR_α. We update this state variable recursively by

$$\phi_n = \phi_{n-1} - \gamma_n L(\phi_{n-1}, U_n), \ n \geq 1, \tag{5.87}$$

where $(U_n)_{n \geq 1}$ is an i.i.d. sequence with distribution U (and probability density p) and for $i = 2, \cdots, d+1$,

$$L_1(\xi, \theta, \mu_1, u)) = e^{-\rho|\mu_1|^b} \left(1 - \frac{1}{1-\alpha} \mathbf{1}_{\left\{L^{(+\mu_1)} - \theta . X^{(+\mu_1)} \geq \xi\right\}} \frac{p(u+\mu_1)}{p(u)}\right), \tag{5.88}$$

$$L_i(\xi, \theta, \mu_i, u)) = \frac{e^{-\rho|\mu_i|^b}}{\left(1 + \tilde{G}^{2c}(-\mu_i)\right)^{1/2}} X^{(+\mu_i)} \mathbf{1}_{\left\{L^{(+\mu_i)} - \theta . X^{(+\mu_i)} \geq \xi\right\}} \frac{p(u+\mu_i)}{p(u)}, \tag{5.89}$$

$$L_{d+2}(\xi, \theta, C, \mu_{d+2}, u) = C - \xi - \frac{1}{1-\alpha} \left(L^{(+\mu_{d+2})} - \theta . X^{(+\mu_{d+2})} - \xi\right)_+ \frac{p(u+\mu_{d+2})}{p(u)}, \tag{5.90}$$

$$L_{d+2+j}(\xi, \theta, \mu_j, u) = K_j(\mu_j, \xi, \theta, u), \quad j = 1, \cdots, d+2, \tag{5.91}$$

The following proposition establishes the *a.s.* convergence of the procedure.

Proposition 5.4.1. *Suppose that the assumptions of Theorem 5.3.3 holds true. Moreover, assume that* (5.70), (5.71), (5.75), (5.76), (5.77) *and* (5.82) *are satisfied. Now if the step sequence* $(\gamma_n)_{n\geq 1}$ *satisfies* (5.45), *then*

$$\phi_n \xrightarrow{a.s.} \phi_\alpha^* := \left(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*, \mu_{1,\alpha}^*, \cdots, \mu_{d+2,\alpha}^*\right).$$

The proof can be deduced by a straightforward adaptation of the proof of Proposition 3.1 in [6].

Now, we are interested by the rate of convergence of the procedure. It shows that the algorithm behaves as expected under quite standard assumptions: it satisfies a Gaussian CLT with optimal rate and minimal variances.

Theorem 5.4.2. *Suppose that the assumptions of Theorem 5.3.6 holds true. Moreover, assume that* (5.70), (5.71), (5.75), (5.76), (5.77) *and* (5.82) *are satisfied. Suppose that the step sequence is defined by* $\gamma_n = \frac{\gamma_1}{n^\beta}$, *with* $\frac{1}{2} < \beta < 1$ *and* $\gamma_1 > 0$. *Let* $\left(\overline{\xi}_n, \overline{\theta}_n, \overline{C}_n\right)_{n\geq 1}$ *be the sequence of Cesàro means defined by:*

$$\overline{\xi}_n := \frac{\xi_0 + \cdots + \xi_{n-1}}{n}, \quad \overline{\theta}_n := \frac{\theta_0 + \cdots + \theta_{n-1}}{n}, \quad \overline{C}_n := \frac{C_0 + \cdots + C_{n-1}}{n}, \ n \geq 1,$$

where (ξ_n, θ_n, C_n) *are defined by* (5.87).

This sequence satisfies the following CLT:

$$\sqrt{n}\begin{pmatrix} \overline{\xi}_n - \xi_\alpha^* \\ \overline{\theta}_n - \theta_\alpha^* \\ \overline{C}_n - C_\alpha^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^*) \qquad as\ n \to +\infty, \tag{5.92}$$

where, for $i = 1, \cdots, d+1$,

$$\Sigma_{1,1}^* = \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^2}\Big(\left(1 + V^T\Pi^{-1}V\right)^2 \times$$
$$\mathrm{Var}\left(\mathbf{1}_{\left\{L^{(+\mu_{1,\alpha}^*)} - \theta_\alpha^*.X^{(+\mu_{1,\alpha}^*)} \geq \xi_\alpha^*\right\}}\frac{p(U+\mu_{1,\alpha}^*)}{p(U)}\right) + \Pi^{-1}V^T\Lambda^*\Pi^{-1}V\Big)$$

$$\Sigma_{i,i}^* = \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^2}\left(m_i^2 \times\right.$$
$$\left.\mathrm{Var}\left(\mathbf{1}_{\left\{L^{(+\mu_{1,\alpha}^*)} - \theta_\alpha^*.X^{(+\mu_{1,\alpha}^*)} \geq \xi_\alpha^*\right\}}\frac{p(U+\mu_{1,\alpha}^*)}{p(U)}\right) + \tilde{m}_{.i}^T\Lambda^*\tilde{m}_{.i}\right)$$

$$\Sigma_{d+2,d+2}^* = \left(\frac{1}{1-\alpha}\right)^2 \mathrm{Var}\left(\left(L^{(+\mu_{d+2})} - \theta_\alpha^*.X^{(+\mu_{d+2,\alpha}^*)} - \xi_\alpha^*\right)_+ \frac{p(U+\mu_{d+2,\alpha}^*)}{p(U)}\right),$$

where $\Lambda^* = \left(\Lambda_{i,j}^*\right)_{1\leq i,j\leq d}$ *and for* $i, j = 1, \cdots, d$,

$$\Lambda_{i,j}^* = \mathbb{E}\left[X_i^{(+\mu_{i,\alpha}^*)}\mathbf{1}_{\left\{L^{(+\mu_{i,\alpha}^*)} - \theta_\alpha^*.X^{(+\mu_{i,\alpha}^*)} \geq \xi_\alpha^*\right\}} \times\right.$$
$$\left. X_j^{(+\mu_{j,\alpha}^*)}\mathbf{1}_{\left\{L^{(+\mu_{j,\alpha}^*)} - \theta_\alpha^*.X^{(+\mu_{j,\alpha}^*)} \geq \xi_\alpha^*\right\}}\frac{p(U+\mu_{i,\alpha}^*)p(U+\mu_{j,\alpha}^*)}{p^2(U)}\right].$$

Remark 6. • *There exists a CLT for the whole sequence* $(\phi_n)_{n\geq 1}$ *and for its empirical mean* $(\overline{\phi}_n)_{n\geq 1}$ *according to Ruppert and Polyak averaging principle. We only stated the result for the three components of interest* $(\overline{\xi}_n, \overline{\theta}_n, \overline{C}_n)$ *since we only need rough estimates for the IS parameters.*

• *The same problem as the one noticed in the averaged global algorithm proposed in [6] to estimate both VaR and CVaR is raised by the algorithm* (5.86)*. Indeed, we added IS procedures to* $(\xi_n, \theta_n, C_n)_{n\geq 1}$ *to improve the convergence toward* $(\xi^*_\alpha, \theta^*_\alpha, C^*_\alpha)$ *but the adjusment of the parameters* μ_i*,* $i = 1, \cdots, d+2$ *need some samples* U_{n+1} *satisfying* $F(U_{n+1} - \mu_{i,n}) - \theta_n . G(U_{n+1} - \mu_{i,n}) > \xi_n$ *which tend to become rare events. We postponed the problems resulting from rare events on the IS procedures itself which may freeze. One may introduce a* companion *VaR procedure (without IS) that will lead the IS parameters to the critical risk area during a first phase of the algorithm which will depend on a moving confidence level* α_n*. Numerical results show that it speeds up the initialzation phase and improve the reduction of variance. For more details, we refer to [6].*

• *In practice we divide our procedure into two phases: Phase I is devoted to the estimation of the variance reducers* $\mu^*_{i,\alpha}$*,* $i = 1, \cdots, d+2$ *using* (5.86) *and a moving confidence level (when* α *is close to 1) with* M *iterations (*M *is small compared to the total number of iterations* N*); Phase II produces some estimates for* $(\xi^*_\alpha, \theta^*_\alpha, C^*_\alpha)$ *based on the procedure defined by* (5.87) *and its Cesàro mean with* $N - M$ *iterations. During this phase, one can either keep updating the IS parameters adaptively or only update* (ξ_n, θ_n, C_n) *using* (5.87) *with frozen IS parameters at* $\hat{\mu}_{i,M}$*,* $i = 1, \cdots, d+2$*. As concerns practical implementation, our numerical results shows that this last choice produces better estimates.*

5.4.2 Linear Control Variable

LCV, which is among the most widely used variance reduction techniques, takes advantage of known properties of simulated model to improve the accuracy of the estimates. It relies on knowing the expectation of simulated random variables which is compared with the estimated expectation obtained during simulation.

In our framework, a natural LCV is $X = G(U)$ since Assumption 5.1.1 implies $\mathbb{E}[X] = 0$. However, if there is more information about the model (quantile of X or L, expectation of L, ...), one can include other LCV to achieve better variance reduction.

We slightly modify the function H of the original algorithm (5.50) in order to introduce LCV, namely

$$\begin{aligned}
\tilde{H}_1(\phi, u) &= H_1(\phi, u) + \beta_1^T X, \\
\tilde{H}_{2:d+1}(\phi, u) &= H_{2:d+1}(\phi, u) + X_\beta, \\
\tilde{H}_{d+2}(\phi, u) &= H_{d+2}(\phi, u) + \beta_{d+2}^T X,
\end{aligned}$$

where $\beta_1, \beta_{d+2} \in \mathbb{R}^d$ and $\Lambda_\beta = (\beta_2 X, \cdots, \beta_{d+1} X)^T$, $\beta_i \in \mathbb{R}$, $i = 2, \cdots, d+1$. Note that $\mathbb{E}\left[\tilde{H}(\phi, U)\right] = h(\phi)$, with $\tilde{H}(\phi, U) = \left(\tilde{H}_1(\phi, U), \tilde{H}_{2:d+1}(\phi, U), \tilde{H}_{d+2}(\phi, U)\right)$, so that the matrix P in Theorem 2.7 is not modified. Actually, only the matrix Γ is

modified and is replaced by $\tilde{\Gamma}$, which is given by

$$\begin{aligned}
\tilde{\Gamma}_{1,1} &= \frac{\alpha}{1-\alpha} + \mathbb{E}\left[\left(\beta_1^T X\right)^2\right], \\
\tilde{\Gamma}_{1,i} &= \tilde{\Gamma}_{i,1} = \mathbb{E}\left[\beta_1^T X X_\beta^T\right], \\
\tilde{\Gamma}_{1,d+2} &= \tilde{\Gamma}_{d+2,1} = \Gamma_{1,d+2} + \frac{1}{1-\alpha}\mathbb{E}\left[\beta_1^T X \left(L - \theta_\alpha^*.X - \xi^*\right)_+\right] + \mathbb{E}\left[\left(\beta_1^T X\right)\beta_{d+2}^T X\right], \\
\tilde{\Gamma}_{i,j} &= \tilde{\Gamma}_{j,i} = \frac{1}{(1-\alpha)^2}\mathbb{E}\left[\left(X_{i-1}\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}} - (1-\alpha)\beta_i X_{i-1}\right)\right. \\
&\qquad\qquad \left.\times\left(X_{j-1}\mathbf{1}_{\{L-\theta_\alpha^* X\geq\xi_\alpha^*\}} - (1-\alpha)\beta_j X_{j-1}\right)\right], \\
\tilde{\Gamma}_{i,d+2} &= \tilde{\Gamma}_{d+2,i} = \Gamma_{d+2,i} + \frac{1}{1-\alpha}\left(\mathbb{E}\left[X_\beta\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+\right] + \mathbb{E}\left[\beta_{d+2}^T X X \mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]\right) \\
&\qquad\qquad + \mathbb{E}\left[X_\beta \beta_{d+2}^T X\right], \\
\tilde{\Gamma}_{d+2,d+2} &= \frac{1}{(1-\alpha)^2}\mathrm{Var}\left(\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+ - (1-\alpha)\beta_{d+2}^T X\right).
\end{aligned}$$

The matrix $\tilde{\Gamma}$ depends deeply of the control variates used: the simple form of the terms $\tilde{\Gamma}_{1,i}$, $i = 2, \cdots, d+2$, is due to the choice of X as a control variate and to the fact that $\mathbb{E}\left[X\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right] = 0$. Consequently, the matrix may be more complex if one chooses other control variates. Now, for sake of simplicity and in order to obtain a matrix $\tilde{\Gamma}$ similar to Γ, *i.e.* $\Gamma_{1,i} = 0$, $i = 2, \cdots, d+1$, we set $\beta_1 \equiv 0$, so that the new asymptotic covariance matrix $\tilde{\Sigma}$

$$\begin{aligned}
\tilde{\Sigma}_{1,1} = \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^2}\Big(&\left(1 + V^T\Pi^{-1}V\right)^2\alpha(1-\alpha) \\
&+ \left(\Pi^{-1}V\right)^T\mathbb{E}\left[\left(X\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}} - (1-\alpha)X_\beta\right)\right. \\
&\left.\left.\times\left(X\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}} - (1-\alpha)X_\beta\right)^T\right]\Pi^{-1}V\Big),
\end{aligned} \tag{5.93}$$

$$\begin{aligned}
\tilde{\Sigma}_{i,i} = \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^2}\Big(&m_i^2\alpha\left(1-\alpha\right) \\
&+\tilde{m}_{.i}^T\mathbb{E}\left[\left(X\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}} - (1-\alpha)X_\beta\right)\left(X\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}} - (1-\alpha)X_\beta\right)^T\right]\tilde{m}_{.i}\Big),
\end{aligned} \tag{5.94}$$

$$\tilde{\Sigma}_{d+2,d+2} = \frac{1}{(1-\alpha)^2}\mathrm{Var}\left(\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+ - (1-\alpha)\beta_{d+2}^T X\right). \tag{5.95}$$

Now, in order to improve the convergence of the original algorithm, we follow the idea developed in the beginning of the last paragraph, *i.e.* we plan to minimize for $i = 1, \cdots, d$,

$$\mathbb{E}\left[\left(X_i\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}} - (1-\alpha)\beta_{i+1}X_i\right)^2\right], \quad \text{and } \mathrm{Var}\left(\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+ - (1-\alpha)\beta_{d+2}^T X\right).$$

The minimum is found by setting each gradient to zero and the optimal least square variables are given by

$$\beta_{i+1}^* = \frac{\mathbb{E}\left[X_i^2\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]}{(1-\alpha)\,\mathbb{E}\left[X_i^2\right]}, \ i = 1, \cdots, d \ \text{ and } \beta_{d+2}^* = \frac{1}{1-\alpha}\Sigma_X^{-1}\Sigma_{XY},$$

where $\Sigma_X := \mathbb{E}\left[XX^T\right]$ and $\Sigma_{XY} := \mathbb{E}\left[X\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+\right]$. This quantities are estimated adaptively during the RM algorithm using the same *innovations*. Moreover, it is always possible to combine the IS algorithm with LCV in order to achieve better variance reduction.

Remark 7. *In this paragraph, we proposed a LCV method based on X to accelerate the convergence of the estimate of $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$. However, due to the particular form of the matrix M one may use another control variable Ξ with known expectation (without loss of generality we can suppose $\mathbb{E}[\Xi] = 0$) only in order to improve the convergence of the sequence $(C_n)_{n\geq 1}$ toward C_α^*. The optimal coefficient β_{d+2}^* is given by*

$$\beta_{d+2}^* = \frac{1}{1-\alpha}\Sigma_\Lambda^{-1}\Sigma_{\Lambda Y},$$

where $\Sigma_\Xi := \mathbb{E}\left[\Xi\Xi^T\right]$ and $\Sigma_{\Xi Y} := \mathbb{E}\left[\Xi\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+\right]$. Substituing this value in the asymptotic variance of the sequence $(C_n)_{n\geq 1}$ we find

$$\mathrm{Var}\left(\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+ - (1-\alpha)\left(\beta_{d+2}^*\right)^T X\right) = \mathrm{Var}\left(\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+\right)\left(1 - R^2\right),$$

where

$$R^2 = \frac{1}{\mathrm{Var}\left(\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+\right)}\Sigma_{\Xi Y}^T\Sigma_\Xi^{-1}\Sigma_{\Xi Y}.$$

5.5 Numerical Examples

5.5.1 Static setting

First we consider two simple examples in the static framework in order to show the efficiency of the CVaR hedging algorithm and of the two variance reduction techniques. For all example, we use RM algorithm with two phases (see Remark 4) combined with the Ruppert & Polyak's averaging principle. In all examples, we define the step sequence by $\gamma_n = \frac{1}{n^\beta}$, with $\beta = \frac{3}{4}$.

Spark Spread

We consider a short position on an exchange option between gas and electricity (called spark spread). Since Electricity has very limited storage possibilities, the seller of this option hedges by trading only gas spot contracts. The process Z can be considered as the electricity spot price since it is observable on the energy market but cannot be used to set up hedging strategies. We choose to model the price of the two spot contracts by the Black & Scholes model with a correlation $\rho = 0.8$ between the two Brownian motions. The loss L can be written

$$L = \left(S_T^e - h_R S_T^g - C\right)_+,$$

where the time horizon $T = 1$ (year), the heat rate $h_R = 4$BTU/kWh (BTU: British Thermal Unit), the generation costs $C = 3$\$/MWh, the two volatilities $\sigma_g = 0.4$, $\sigma_e = 0.8$ and the electricity and gas initial spot prices are $S_0^e = 40$\$/MWh, $S_0^g = 3$\$/MMBTU. The seller of the option uses a self-financed static strategy based on the gas spot price in order to reduce its risk at time $t_0 = 0$. Thus, its optimal strategy is given by the solution of (5.3) with $\ell = 0$. A crude Monte Carlo gives $\mathbb{E}[L] = 11.86$

	No hedging		Static hedging		
α	VaR	CVaR	VaR	θ^*_α	CVaR
95%	65.1	114.4	63.1	7.8	98.3
99%	142.2	208.3	120.2	13.6	163.2
99.5%	183.1	257.8	146.8	16.4	190.2

Table 5.1: One step Self-financed static CVaR-hedging of Spark Spread option

α	$\mathrm{VR}_{\mathrm{VaR}}$(IS)	$\mathrm{VR}_{\mathrm{Reg}}$ (IS)	$\mathrm{VR}_{\mathrm{CVaR}}$ (IS)	$\mathrm{VR}_{\mathrm{CVaR}}$(LCV)
95%	3.0	1.9	16.7	2.0
99%	3.7	2.3	19.0	1.7
99.5%	4.5	3.0	20.2	1.5

Table 5.2: Variance Ratios for the One-step Self-financed static CVaR-hedging of Spark Spread option

with a variance of 3692 after 3 000 000 trials. The variance ratios correspond to an estimate of the asymptotic variance obtained without any variance reduction techniques, *i.e.* (5.56), (5.57) and (5.58) divided by an estimate of the asymptotic variance using IS (column (IS)) or LCV (column (LCV)) (see the asymptotic matrix obtained in Section 5.4).

In this example, the LCV method based on X doesn't provide any variance reduction. However, for the CVaR component, we use the control

$$\Lambda = \mathbf{1}_{\left\{S_T^e \geq q_\delta^e\right\}} - (1-\delta),$$

where q_δ^e is the quantile of S_T^e at level δ. We choose: $\delta = 0.995$ ($q_\delta^e \approx 228.04$) for $\alpha = 0.95$, $\delta = 0.999$ ($q_\delta^e \approx 344.15$) for $\alpha = 0.99$ and $\delta = 0.9995$ ($q_\delta^e \approx 403.95$) for $\alpha = 0.9995$. The results obtained for three different values of the confidence level $\alpha = 95\%,\ 99\%,\ 99.5\%$ after 3 000 000 iterations of the Robbins-Monro procedure are specified in Table 1. We provide the VaR and CVaR of the loss without any hedging strategy which are computed using the Robbins-Monro procedure developed in [6].

To complete this numerical example, we provide the histograms of the loss obtained with and without hedging. We clearly see on Figure 5.1 that the asymetry of the histogram has been changed from right(loss) to left (gain) so that gains are more likely to occured with the hedged portfolio. In order to change the right tail distribution of the loss, the mode of the original portfolio has been greatly reduced and slightly translated to the right. Figure 5.2 confirms this idea: in order to hedge rare events that happen in the right tail distribution, the strategy consists in enlarging the left tail distribution. This induces a slight reduction of the mode and its translation to the right.

Consumption hedging

At time $T = 1$ (year), an energy provider buys on an energy market a quantity C_T of gas at price S_T^g and sells it to consumers at a fixed price $K = 11€/\mathrm{MWh}$. The quantity C_T denotes the consumption at time T and is equal to $C_T = a - bT_T$, with

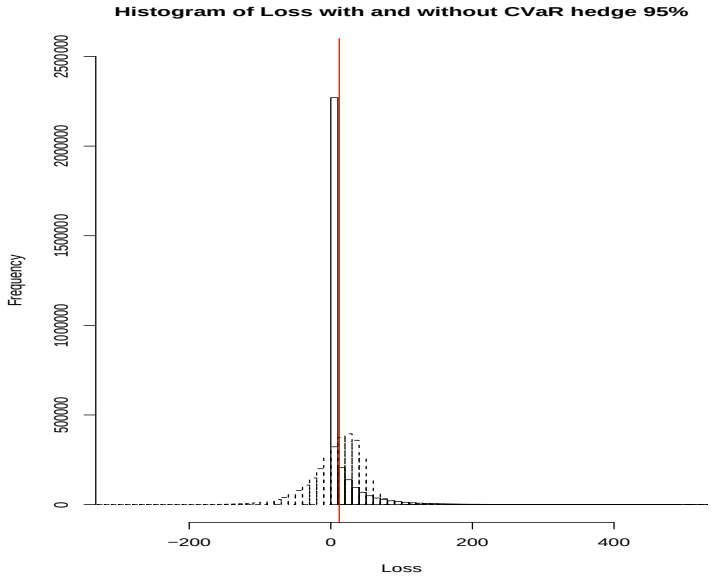


Figure 5.1: Histogram of loss with (dashed lines) and without (normal lines) one step CVaR-hedging at level $\alpha = 95\%$. The vertical line is the mean of the portfolio loss distribution.

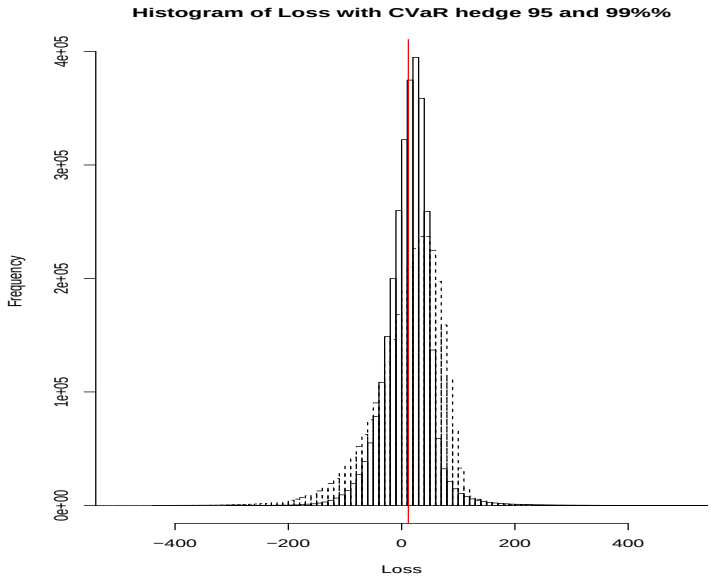


Figure 5.2: Histogram of one step CVaR-hedged loss at level $\alpha = 95\%$ (normal lines) and $\alpha = 99\%$ (dashed lines). The vertical line is the mean of the portfolio loss distribution.

$a = 100$ Mwh and $b = 3$ MWh/°C. The temperature is modeled as an Ornstein-Uhlenbeck process so that the temperature at time T is given by

$$T_T = e^{-\lambda T} T_0 + m(1 - e^{-\lambda T}) + \sigma_T \sqrt{\frac{1 - e^{-2\lambda T}}{2\lambda}} G_1,$$

with $T_0 = 11°\text{C}$, $\lambda = 0.02$, $m = 11°\text{C}$, $\sigma_T = 6°\text{C}$ and $G_1 \sim \mathcal{N}(0,1)$. Gas spot price is modeled as a geometric Brownian motion with $S_0 = 11$ €/MWh and the Brownian motion of gas spot price is correlated with the one of the temperature, $\rho = -0.8$, namely

$$S_T = S_0 e^{-\frac{\sigma_g^2}{2}T + \sigma_g\sqrt{T}\left(\rho G_1 + \sqrt{1-\rho^2}G_2\right)},$$

where $\sigma_g = 0.4$, $G_2 \sim \mathcal{N}(0,1)$, and is independent of G_1. Consequently, the loss suffered by the energy provider at time T is given by

$$L = (S_T - K)C_T.$$

The energy provider uses a self-financed static strategy based on the gas spot price in order to reduce its risk at time $t_0 = 0$. A crude Monte Carlo gives $\mathbb{E}[L] = 62.6$ with a variance of 10747.4 after 3 000 000 trials.

In this example, the functions $Q_i(.,\xi_\alpha^*,\theta_\alpha^*)$, $i = 1,\cdots,d+2$ defined by (5.78), (5.79) and (5.80) reach their minimum for $\mu_{i,\alpha}^* \approx 0$, so that the IS algorithms doesn't achieve any significative variance reduction. Moreover, the LCV method based on X doesn't achieve a significative variance reduction. Consequently, for the hedged portfolio, we don't provide any variance reduction ratio. However, we notice that in order to estimate both VaR and CVaR of the loss L without hedging portfolio, the IS algorithm provides significant variance reduction. Those results are due to the fact that, in this example, CVaR hedging already appears as a way to reduce the variance of the loss. Consequently, reducing again the different variances by IS doesn't provide any further variance reduction whereas in the first example, CVaR hedging did not reduce the variance of the original loss but tries to capture some gains in order to reduce the global CVaR so that IS and LCV succeeds in reducing the considered variances. Results are summarized in Table 5.3.

To complete this numerical example, we provide the histograms of the loss obtained with and without CVaR hedging using 3 000 000 samples. We can see on Figure 5.3. that the right tail distribution (which corresponds to high loss) is greatly reduced. The deformation provided by a CVaR hedging at level 95% is very impresive. The mode of the hedged loss distribution has been translated to the right near 0 whereas without hedging it was negative, which means that the loss occuring the most frequently has changed from negative (gain) to positive value (loss). In order to reduce the right heavy tail which corresponds to high loss, the CVaR hedging strategy translates the mode near the mean and thus gives more probability to small losses. Figure 5.4. illustrates the histrograms obtained with a CVaR hedging at level 95% and 99%. We remark that the distribution which corresponds to a CVaR hedging at level 99% has heavier tails than the one corresponding to a CVaR hedging at level 95%. The more α is close to 1, the heavier CVaR-hedged loss distribution tails are. Note that the mode of the distribution slightly translated to the left.

5.5.2 Dynamic setting

We keep on studying the consumption hedging example and now, we experiment our 4 different algorithms to compute the optimal self-financed dynamic strategy: C.H., B.H., M.D.H. and C.D.H. (see Section 3 for more details about each strategy and the RM algorithm associated). The parameters of the last example remain unchanged.

	No hedging		Static hedging		
α	VaR	CVaR	VaR	θ^*_α	CVaR
95%	784.6	1226.3	259.6	81.6	366.5
99%	1452.4	2012.3	437.1	89.9	537.3
99.5%	1769.9	2382.8	505.7	92.3	608.6

Table 5.3: Self-financed static hedging of Consumption

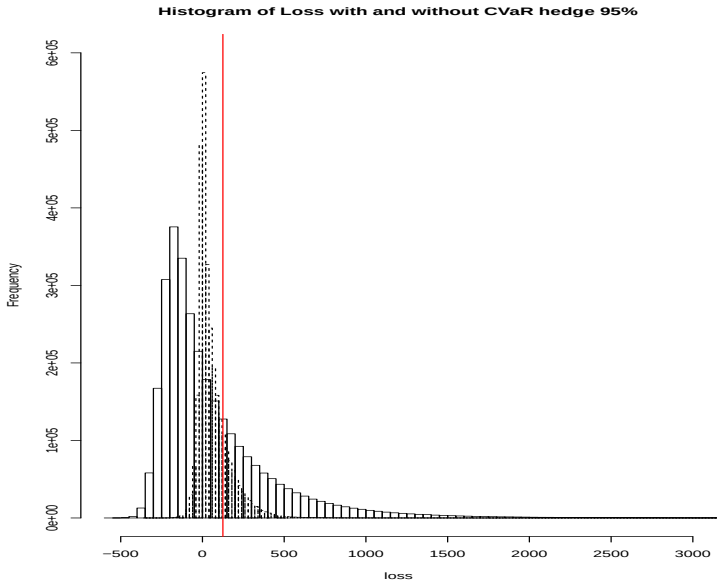


Figure 5.3: Histogram of loss with (dashed lines) and without (normal lines) one step CVaR-hedging at level $\alpha = 95\%$. The vertical line is the mean of the portfolio loss distribution.

We consider 3 different values for the number of trading dates: $M = 4$ (one trade each trimester), $M = 12$ (one trade each month), $M = 52$ (one trade each week) and the CVaR-hedging level is 95%. All layers in the quantization tree of the process $(X_\ell, Z_\ell)_{1\leq \ell \leq M}$ have the same size, *i.e.* $N = N_\ell = 10$, $\ell = 1, \cdots, M$. Note that we do not quantify the process $(S_{t_\ell}, T_{t_\ell})_{1\leq M-1}$ but only the two gaussian random variables $(G1, G2)$ so that our quantization trees are obtained as a transform of the 2-dimensional normal distribution optimal grid. It is crucial to have a good approximate of the random variable $\tilde{\Delta} L_\ell$, $\ell = 1, \cdots, M$, for the method M.D.H. so that we use an optimized quantization grid of size 100 in (5.67). Results are summarized in Table 3.

We clearly see that the optimal strategy is given by the M.D.H. method when the number of trading dates becomes large. The C.H. method for $M \leq 12$ is optimal but suffers from convergence when $M \geq 12$. When M is large enough, the dimension of the algorithm in the C.H. method becomes too high and the estimate of the optimal strategy doesn't converge anymore. The larger is the number of trading dates, the greater is the difference between the M.D.H. and C.D.H. methods. Figure 5.5 presents the histograms of the loss without any hedging strategy and with a

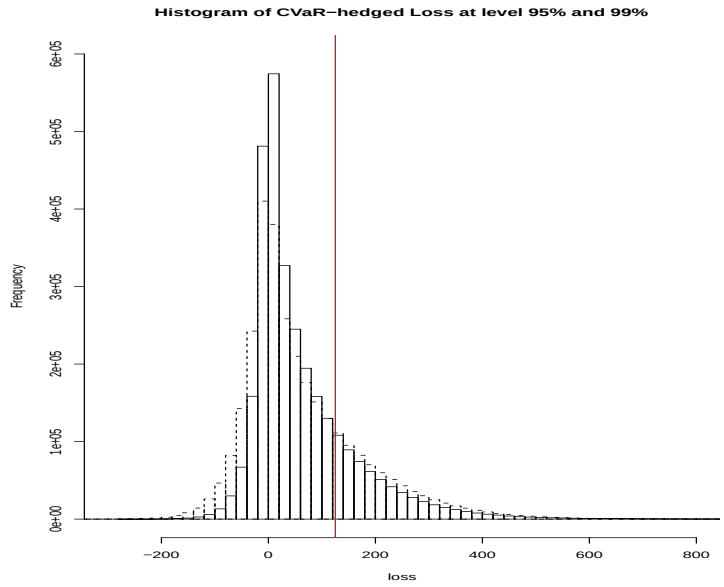


Figure 5.4: Histogram of one step CVaR-hedged loss at level $\alpha = 95\%$ (normal lines) and $\alpha = 99\%$ (dashed lines).

method	C.H.		B.H.		M.D.H.		C.D.H.	
M	VaR	CVaR	VaR	CVaR	VaR	CVaR	VaR	CVaR
4	178.3	240.9	175.9	252.5	177.8	252.9	178.9	259.2
12	163.2	214.1	160.7	233.8	158.7	221.7	161.9	232.9
52	272.6	395.1	158	233.2	148.7	210.1	153.1	223.7

Table 5.4: Self-financed dynamic CVaR hedging of Consumption at level 95% with different strategies.

CVaR-hedging at level 95% using the M.D.H. method with 52 trading dates. The deformation of the loss distribution is very impresive. Like in the static framework, the mode of the CVaR hedged loss distribution has been translated near the mean and in order to reduce the right tail distribution, the CVaR hedging strategy makes middle loss more likely. Figure 5.6 compares the CVaR hedged loss distribution at level 95% using the static strategy and the dynamic strategy M.D.H. with 52 dates. The dynamic strategy translates the mode on the mean and removes losses under the mean to reduce the right tail distribution. Note that the very left tail of the two distributions (which corresponds to gains) are quite similar: dynamic strategy reduces greatly high losses and slightly high gains. Figure 5.7 shows the 10 components of the optimal trading strategy using the M.D.H with 52 trading dates.

5.6 Appendix: Proof of Proposition 5.2.5

We propose below the proof of (5.21) and (5.22) which are the key results in order to derive our R.M. algorithm. The proof of (5.19) and (5.20) will follow using similar arguments.

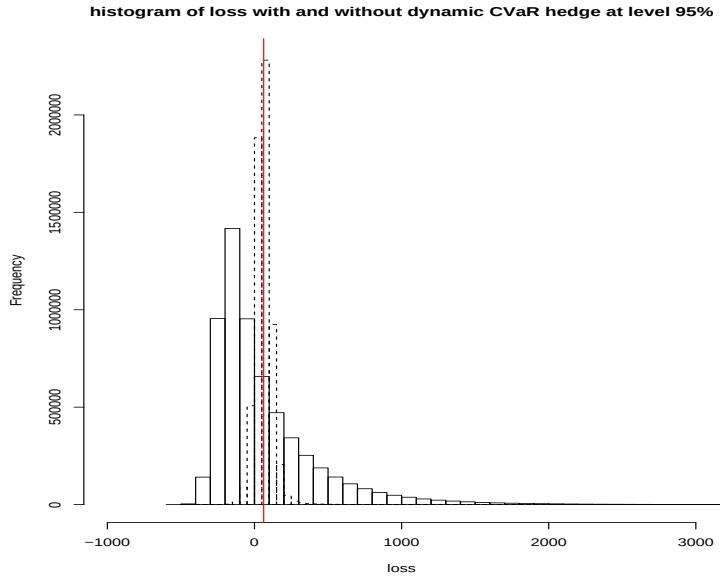


Figure 5.5: Histogram of Consumption's Loss with (dashed lines) and without (normal lines) dynamic CVaR-hedging at level $\alpha = 95\%$ using the M.D.H. strategy (52 trading dates).

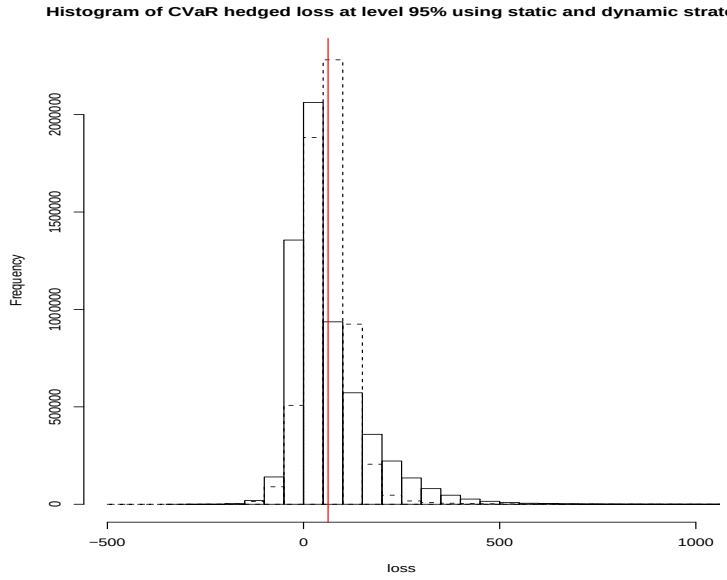


Figure 5.6: Histogram of CVaR hedged Loss at level $\alpha = 95\%$ with one step static (normal lines) and dynamic (dashed lines, M.D.H. with 52 trading dates) self-financed strategies.

Proof. First note that since $L \in L^1_{\mathbb{R}}(\mathbb{P})$, $\mathbb{E}[L|\mathcal{F}] \in L^0_{\mathbb{R}}(\mathcal{F})$ so that

$$\mathbb{E}[\mathcal{F}\text{-CVaR}_\alpha(L - \theta.X)] = \mathbb{E}[L] + \mathbb{E}[\mathcal{F}\text{-CVaR}_\alpha(L - \mathbb{E}[L|\mathcal{F}] - \theta.X)].$$

Consequently, we can suppose that $\mathbb{E}[L|\mathcal{F}] = 0$ for the rest of the proof. It is

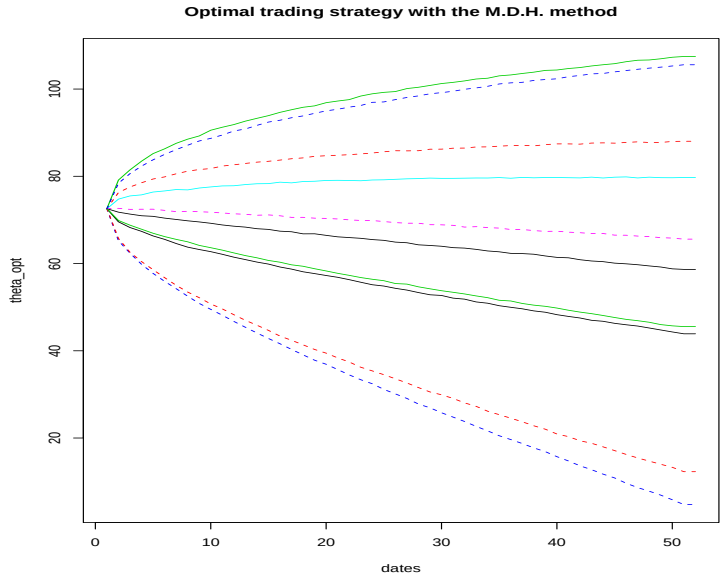


Figure 5.7: Optimal trading strategy using the M.D.H. at level $\alpha = 95\%$ with 52 trading dates and 10 quantization nodes per dates.

straightforward that

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right] \geq \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F}), \xi \in L^0_{\mathbb{R}}(\mathcal{F})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ |\mathcal{F}\right]\right].$$

Let $(\theta_n)_{n\geq 1}$ be a sequence in $L^0_{\mathbb{R}^d}(\mathcal{F})$ such that

$$\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F}), \xi \in L^0_{\mathbb{R}}(\mathcal{F})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ |\mathcal{F}\right] = \inf_{n\geq 1} \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta_n.X\right),$$

and consider the sequence $(\Xi_n)_{n\geq 1}$ with $\Xi_1 = \theta_0 := 0$, and defined recursively for $n \geq 1$ by

$$\Xi_{n+1} := \begin{cases} \Xi_n & \text{, if } \mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi^n.X\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta_n.X\right), \\ \theta_n & \text{, if } \mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi^n.X\right) \geq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta_n.X\right). \end{cases}$$

Note that $\Xi_n \in L^0_{\mathbb{R}^d}(\mathcal{F})$ for $n \geq 1$ and

$$\mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi_{n+1}.X\right) = \min_{0\leq p\leq n} \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta_p.X\right) \quad a.s.,$$

so that the sequence $\left(\mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi_n.X\right)\right)_{n\geq 0}$ is non increasing and by Jensen's inequality

$$\mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi_n.X\right) \geq \frac{1}{1-\alpha}\mathbb{E}\left[L_+|\,\mathcal{F}\right] \geq \frac{1}{1-\alpha}\mathbb{E}\left[L|\,\mathcal{F}\right]_+ = 0.$$

Moreover, by definition for $n \geq 0$

$$\mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi_n.X\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L\right) \leq \frac{1}{1-\alpha}\mathbb{E}\left[L_+ | \mathcal{F}\right] \in L^1_{\mathbb{R}}(\mathbb{P}).$$

The sequence $\left(\mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi_n.X\right)\right)_{n\geq 0}$ converges in $L^1_{\mathbb{R}}(\mathbb{P})$ ought to Beppo-Levi's Theorem and

$$\begin{aligned}\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right] &\geq \inf_{n\geq 0} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi_n.X\right)\right] \\ &\geq \mathbb{E}\left[\inf_n \mathcal{F}\text{-CVaR}_\alpha\left(L - \Xi_n.X\right)\right] \\ &= \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right].\end{aligned}$$

The proof of (5.19) follows using similar arguments.

Let $\omega \in \Omega$. The convexity of $V_f(\omega, ., .)$ is a consequence of the convexity of $(\xi, \theta) \mapsto v_f(\xi, \theta, y, x)$ for all $(y, x) \in \mathbb{R} \times \mathbb{R}^d$. Owing to Jensen's inequality, for all $(\xi, \theta) \in \mathbb{R} \times \mathbb{R}^d$,

$$V_f(\omega, \xi, \theta) \geq \xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi\right)_+.$$

Now, if $\xi < \int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y)$, we have $\xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi\right)_+ = -\frac{\alpha}{1-\alpha}\xi + \frac{1}{1-\alpha}\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) \to +\infty, \xi \to -\infty$. Moreover for all $\xi \in \mathbb{R}$, $\xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi\right)_+ \geq \xi$, thus it imples that $\lim_{\xi\to+\infty} \xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi\right)_+ = +\infty$, which finally yields, for all $\theta \in \mathbb{R}^d$,

$$\lim_{|\xi|\to+\infty} V_f(\omega, \xi, \theta) = +\infty.$$

Now, in order to establish that the function $V_f(\omega, \xi, .)$ goes to infinity as $|\theta|$ goes to infinity for all $\xi \in \mathbb{R}$, we show that $\inf_{\xi\in\mathbb{R}} V_f(\omega, \xi, \theta) = \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)(\omega)$ satisfies

$$\lim_{|\theta|\to+\infty} \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)(\omega) = +\infty.$$

First note that the sub-additivity of the function $x \mapsto x_+$ implies that

$$\mathcal{F}\text{-CVaR}_\alpha\left(-\theta.X\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right) + \mathcal{F}\text{-CVaR}_\alpha\left(-L\right),$$

so that,

$$|\theta|\mathcal{F}\text{-CVaR}_\alpha\left(-\frac{\theta}{|\theta|}.X\right) - \mathcal{F}\text{-CVaR}_\alpha\left(-L\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right),$$

which finally yields,

$$|\theta| \operatorname*{ess\,inf}_{u\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),|u|=1} \mathcal{F}\text{-CVaR}_\alpha\left(u.X\right) - \mathcal{F}\text{-CVaR}_\alpha\left(-L\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right),$$

so that owing to Assumption 5.2.5 *ii*), $\lim_{|\theta|\to+\infty} \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)(\omega) = +\infty$. The proof of (5.20) follows using similar arguments.

Consequently, there exists $(\xi^*_\alpha, \theta^*_{1,\alpha}) := (\xi^*_\alpha(\omega), \theta^*_\alpha(\omega))$ and for all $\xi \in \mathbb{R}$, $\theta^*_{2,\alpha} := \theta^*_\alpha(\omega, \xi)$ which are $\mathcal{F}$-measurable owing to measurable selection theorem (see e.g. Lemma 3 and Lemma 4 in [29]), such that

$$\inf_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi,\theta) = \min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi,\theta) = V_f(\xi^*_\alpha, \theta^*_{1,\alpha}) \quad a.s.$$

and

$$\inf_{\theta\in\mathbb{R}^d} V_f(\xi,\theta) = \min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta) = V_f(\xi, \theta^*_{2,\alpha}) \quad a.s.$$

□

Part III

Joint Modelling of Gas and Electricity spot prices

Chapter 6

Joint Modelling of Gas and Electricity spot prices

Joint paper with V. Lemaire.

Abstract: The recent liberalization of the electricity and gas markets has resulted in the growth of energy exchanges and modelling problems. In this paper, we modelize jointly gas and electricity spot prices using a mean-reverting model which fits the correlations structures for the two commodities. The dynamics are based on Ornstein processes with parameterized diffusion coefficients. Moreover, using the empirical distributions of the spot prices, we derive a class of such parameterized diffusions which captures the most salient statistical properties: stationarity, spikes and heavy-tailed distributions.

The associated calibration procedure is based on standard and efficient statistical tools. We calibrate the model on French market for electricity and on UK market for gas, and then we simulate some trajectories which reproduce well the observed prices behavior. Finally, we illustrate the importance of the correlation structure and of the presence of spikes by measuring the risk on a power plant portfolio.

Keywords: Electricity markets; spot price modelling; ergodic diffusion; stochastic differential equation; saddlepoint.

6.1 Introduction

The recent deregulation of energy markets has led to the development in several countries of market places for energy exchanges. Consequently, understanding and modelling the behavior of energy market is necessary for developing a risk management framework as well as pricing of options. Many derivatives on both electricity and gas spot (and futures) prices are traded. Understanding the correlation structure between both energies is a significant challenge. For instance, spark spread options are commonly traded in energy markets as a way to hedge price differences between electricity and gas prices or are used in order to price projects in energy (see [36] for an introduction). Thus, modelling jointly the evolution of gas and electricity prices is a relevant issue.

Numerous diffusion-type and econometric models have been proposed for electricity and gas spot prices. In energy markets, spot price dynamics are commonly based on Ornstein processes, which are the classical way to model mean-reversion. Geometric models represent the logarithmic prices by a sum of Ornstein processes with different speeds of mean reversion whereas arithmetic models represent the price itself (see for instance [82] for a geometric model). Also, equilibrium models ([7] and [46]) have been investigated in order to reproduce price formation as a balance between supply and demand. The main drawback of such model is that they do not reproduce the autocorrelation structure of a commodity and the cross-correlation structure between commodities. In [37], a markov jump diffusion is investigated for electricity spot prices. Though, it properly represents the spiky behaviour of spot electricity prices, the process reverts to a deterministic mean level whereas it usually reverts to the pre-spike value on data. Moreover applied to electricity and gas spot prices, it does not capture the autocorrelation and cross-correlation struture observed on data.

Another class of spot price dynamics is represented by multifactor models. Several authors (see [10], [20], [66], [85] among others) have investigated this kind of diffusion. The logarithmic prices or the price itself is represented by a sum of Ornstein processes in order to incorporate a mixture of jump variations and "normal" variations. For instance, in [66] the deseasonalized spot price or log-spot price $X(t)$ is given by:

$$X(t) = Y_1(t) + Y_2(t)$$

where

$$\mathrm{d}Y_i(t) = -\lambda_i Y_i(t)\mathrm{d}t + \mathrm{d}L_i(t), \quad i = 1, 2.$$

The Ornstein Uhlenbeck (OU) component Y_1 is responsible for the normal variation and is assumed to be Gaussian, *i.e.* $L_1(t)$ is a Brownian motion, whereas Y_2 is the Levy driven OU component responsible for spikes, *i.e.* $L_2(t)$ is a jump Lévy process. In this kind of framework, the difficulty is to detect and filter the spikes in order to estimate the jump part. Several methods have been proposed to circumvent this problem (see e.g. [66] and [10]). In [12], the following spot price dynamics for

two energies A and B are proposed

$$S^A(t) = \Lambda^A(t) + \sum_{i=1}^{m} X_i^A(t) + \sum_{j=1}^{n} Y_j^A(t),$$
$$S^B(t) = \Lambda^B(t) + \sum_{i=1}^{m} X_i^B(t) + \sum_{j=1}^{n} Y_j^B(t),$$

where $\Lambda^A(t)$ and $\Lambda^B(t)$ are seasonal floors, X_i^A and X_i^B are common OU processes, i.e. they are driven by the same jump process L_i. A different approach based on copula is proposed in [11] where the joint evolution of electricity and gas prices is modeled by a bivariate non-Gaussian OU pure jump process with a non-symmetric copula.

In this paper, we propose an alternative class (arithmetic and geometric) of models to reproduce adequately the statistical features of gas and electricity spot prices based on parameterized local volatility processes. This approach is motivated by [14] where diffusion models with linear drift and prespecified marginal distribution are investigated with an application in a different context. The spiky behaviour of both spot prices is captured without introducing jump diffusion models. Moreover, this approach provides a significant advantage over the class of jump diffusion models since the calibration process involves only classical statistical tools like least squares method so that it is robust and fast. It allows to reproduce (for the first time to our knowledge) both the auto-correlation and the cross-correlation strutures between two energies. The model was successfully tested on several markets and seems to fit well the statistical features and the marginal distributions of gas and electricity spot prices.

Our results are presented as follows. Section 2 is devoted to the description of the stylised features of gas and electricity spot prices. Then, in Section 3, we briefly recall some important theoretical results on which are based our model. To be more precise, we recall how to construct a mean reverting diffusion process X solution of a stochastic differential equation (SDE) with a prespecified continuous invariant density f. Such diffusions involves parameterized local volatility processes. In Section 4, we present the model of our choice and focus on the calibration procedure. In the last section, we perform the calibration on the data sets coming from the NBP for the gas spot price and the Powernext market for the electricity spot price. Then, we proceed to the simulation and, finally, analyze the impact of the modelization by measuring the risk of an energy related portfolio using several models. We show that introducing the cross-commodity correlation structure can greatly modify the risk of a portfolio.

6.2 Stylised features of gas and electricity spot prices

6.2.1 Seasonality

A first characteristic of gas and electricity (and many commodities) prices is the presence of annual (and possibly multi-time scales) seasonality and a trend (see e.g. [37], [66]). For each commodity, we model the seasonality and the trend component

of the logarithmic spot prices with the mean level functions around which spot prices fluctuate

$$\log g(t) = a^g + b^g t + \sum_{k=1}^{m} c_k^g \cos\left(\frac{2\pi t}{l_k}\right) + d_k^g \sin\left(\frac{2\pi t}{l_k}\right),$$
$$\log e(t) = a^e + b^e t + \sum_{k=1}^{m} c_k^e \cos\left(\frac{2\pi t}{l_k}\right) + d_k^e \sin\left(\frac{2\pi t}{l_k}\right),$$

where $l_k = \lfloor 252/k \rfloor$, $k = 1, ..., m$, $\lfloor x \rfloor$ denotes the integer part of x. For instance, if we choose $m = 2$, we only consider a seasonal function over the year and the semester. We assume 252 trading days in a year except for electricity spot price on Powernext which has 365 trading days in a year so that, we have to take into account *this particularity* in the seasonality function. The coefficients above are estimated using ordinary least squares. The log-seasonality functions are represented with the estimated values for $m = 2$ using gas spot price at the NBP and electricity spot price from the Powernext market in Figure 6.1. All parameters are not significant at the 5% level. We only report and take into account the significant values [1]. We checked the seasonality over week, month and quater, but the coefficients were not significant.

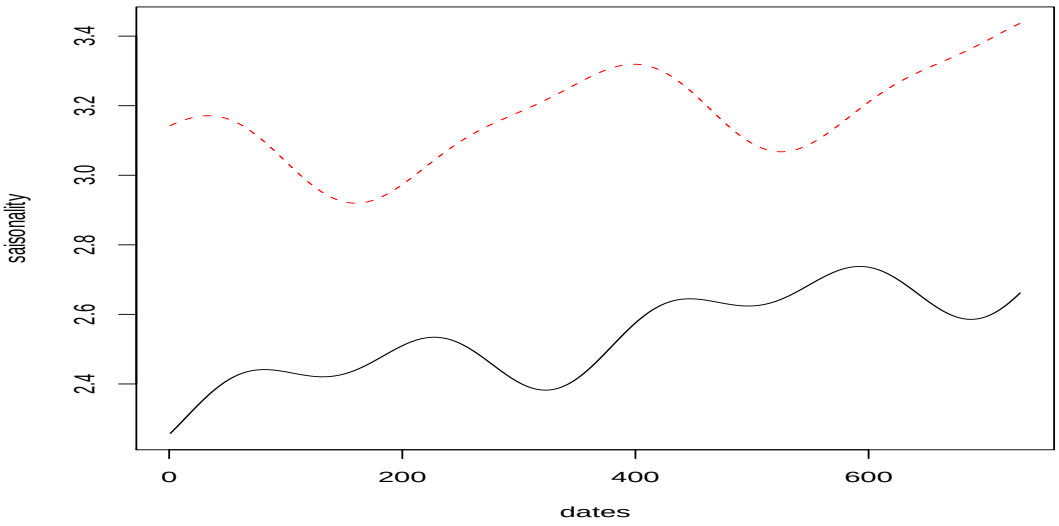


Figure 6.1: The fitted log-seasonality functions $\log(g(t))$ and $\log(e(t))$

Now we focus our attention on the deseasonalized data $Y^g(t) := \log S^g(t) - \log g(t)$ and $Y^e(t) := \log S^e(t) - \log e(t)$ for the specification of the model. A geometric model consists in modelling the stochastic processes $Y^g(t)$ and $Y^e(t)$ whereas an arithmetic model consists in modelling the stochastic processes $e^{Y^g(t)}$ and $e^{Y^e(t)}$.

[1] $a^g = 1.53, b^g = 0.000688, c_1^g = 0.121, d_2^g = 0.0287, c_2^g = 0.00533$ et $a^e = 3.02, b^e = 0.000405, c_1^e = 0.138, d_2^e = 0.0368$.

6.2.2 Spikes and heavy tails

Electricity has very limited storage possibilities. It induces the possibility of spikes in spot prices. Natural gas can be stored but it is often costly, so that it shares the spiky behaviour of spot electricity prices. Gas and electricity markets share this similarity as it can be seen in Figure 6.2 presenting the electricity spot prices coming from the Powernext market on the left and gas spot prices at the National Balancing Point (NBP) on the right. From a stochastic modelling point of view, spikes are commonly represented by jump diffusions with mean reversion. However (to the best of our knowledge) there is no evidence that it is rather jumps than spikes caused by clusters of volatility for instance.

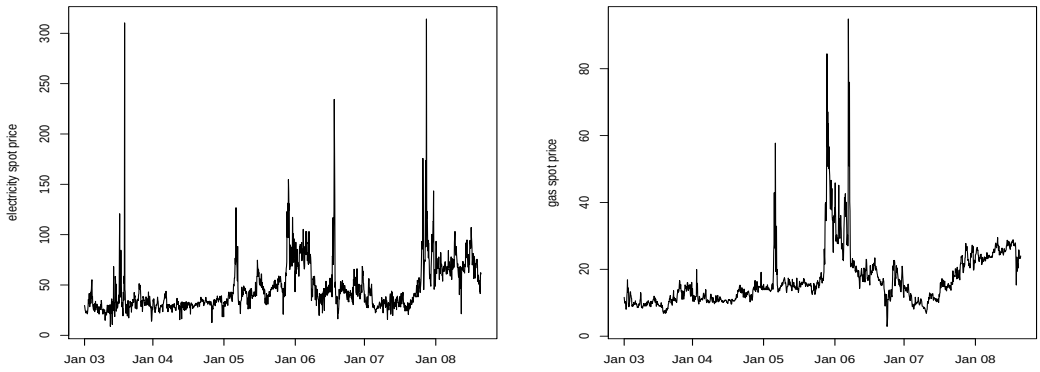


Figure 6.2: Electricity spot prices on the Powernext market (on the left) and gas spot prices at the NBP (on the right) for the period 14 January 2003 till 20 August 2008.

The histograms of Y^g and Y^e with the fitted normal density curve is presented in Figure 6.3. We observe that the two residuals time series Y^g and Y^e are far from being normally distributed. The excess of kurtosis of Y^g and Y^e are respectively equal to 4.5 and 2.3 meaning that the two distributions are peaked and have heavy tails. The skewness of Y^g and Y^e are respectively equal to 0.77 and 0.57 meaning that the two distributions are not symmetric.

6.2.3 Mean reversion and long term dependency

Gas and Electricity spot prices are known to be stationary. This can be tested using an augmented Dickey-Fuller test (ADF) or the Phillips-Perron test. For the UKPX, Powernext electricity spot prices and gas spot prices at the NBP the unit root hypothesis was rejected using both tests. Figure 6.4 shows that gas and electricity deseasonalized prices are strongly linked by a long term dependency, *i.e.* it seems that there is a stochastic equilibrium between $Y^g(t)$ and $Y^e(t)$ from which they cannot deviate for a long time. This long term dependency can be observed on the cross-correlation function.

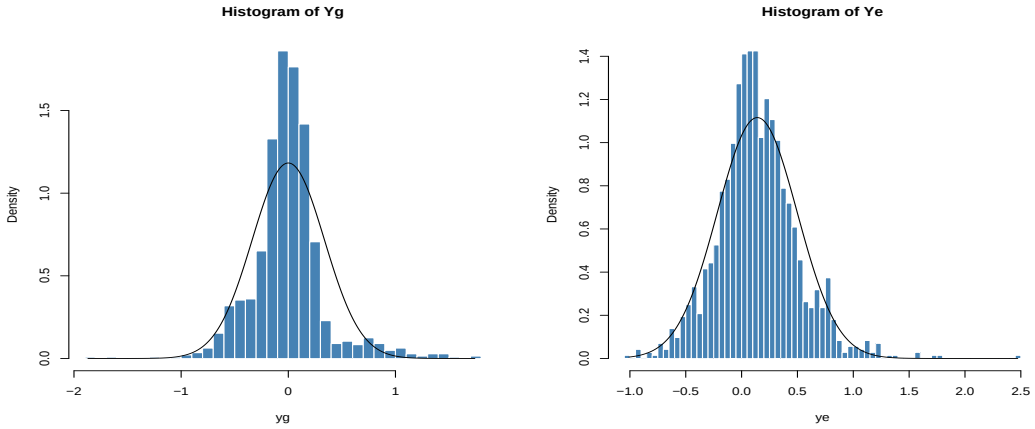


Figure 6.3: Histograms of Y^g and Y^e with normal density curves.

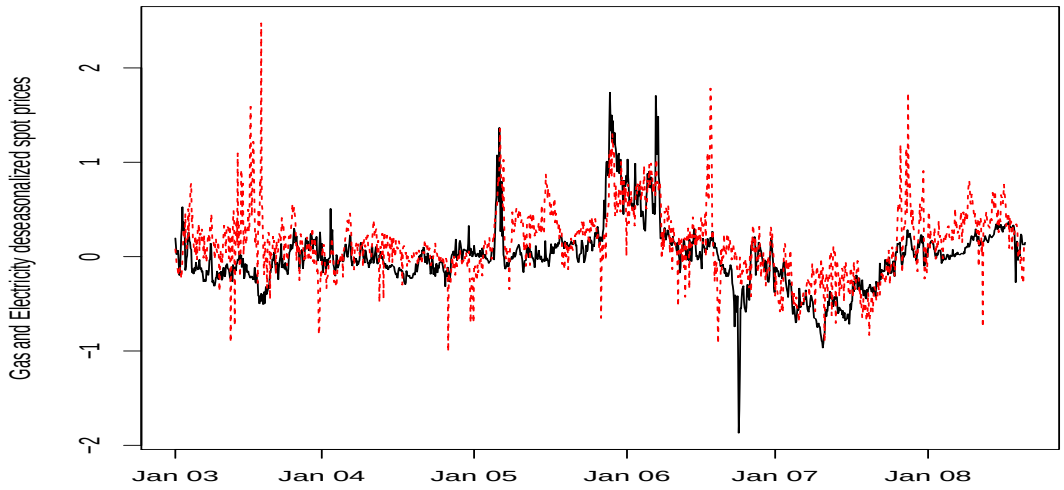


Figure 6.4: The log-deseasonalized gas (normal line) and electricity spot (dashed line) prices

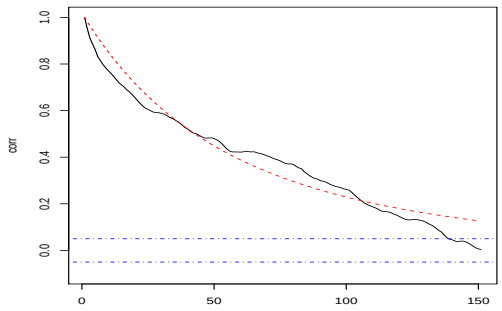
6.2.4 Auto-correlation and cross-correlation

In energy spot price modelling, the auto-correlation functions (ACFs) are often analyzed. The ACFs of both $Y^g(t)$ (respectively $e^{Y^g(t)}$), ρ^g, on one hand $Y^e(t)$ (respectively $e^{Y^e(t)}$), ρ^e, on the other hand present both a two-scale (or three-scale at most) decreasing behaviour with one quickly decreasing component and one or two slow decreasing components. The same behaviour is observed on the cross-correlation function (CCF) $\rho^{g,e}$. This kind of decreasing ACFs and CCF are well

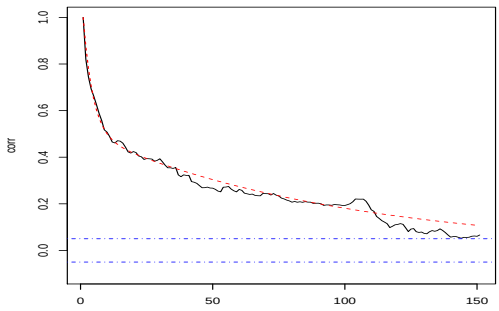
explained by sum of decreasing exponentials components, namely for $\tau > 0$:

$$\rho^g(\tau) = \text{Corr}\left(Y^g(t+\tau), Y^g(t)\right) = \phi_1^g e^{-\lambda_1^g \tau} + (1-\phi_1^g)e^{-\lambda_2^g \tau},$$
$$\rho^e(\tau) = \text{Corr}\left(Y^e(t+\tau), Y^e(t)\right) = \phi_1^e e^{-\lambda_1^e \tau} + (1-\phi_1^e)e^{-\lambda_2^e \tau},$$
$$\rho^{g,e}(\tau) = \text{Corr}\left(Y^g(t+\tau), Y^e(t)\right) = \phi^{g,e} e^{-\lambda^{g,e} \tau}.$$

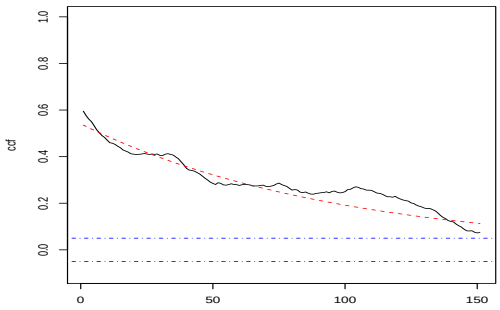
For the sake of simplicity in our stochastic modelization, we focused on one type of cross-correlation Corr $(Y^g(t+\tau), Y^e(t))$ and we assumed that the cross-correlation is symetric that is Corr $(Y^g(t+\tau), Y^e(t)) =$ Corr $(Y^e(t+\tau), Y^g(t))$ which is a rather natural approximation. We observed that the slower rates of mean reversion for each commodities are quite similar $\lambda_2^g = \lambda_2^e$ and that a rather good approximation is obtained by setting $\lambda^{g,e} = \lambda_2^g = \lambda_2^e$. Using a least squares approach, we fitted



(a) ACF of Y^g



(b) ACF of Y^e



(c) CCF of (Y^g, Y^e)

Figure 6.5: Empirical ACF and CCF of deseasonalized gas spot price and electricity spot price

simulteanously $\rho^g(\tau)$, $\rho^e(\tau)$, $\rho^{g,e}(\tau)$ $(\tau = 1, ..., 150)$ to the empirical ACFs and CCF. We assumed that the observed spot prices have reached the stationarity. Both empirical and fitted ACFs[2] and CCF[3] are plotted in Figure 6.5.

[2] $\phi_1^g = 0.43$, $\lambda_1^g = 7.2$, and $\phi_1^e = 0.49$, $\lambda_1^e = 69.4$
[3] $\phi^{g,e} = 0.53$, $\lambda_2^g = \lambda_2^e = \lambda^{g,e} = 2.6$

We can see the separation into a fast speed of mean reversion for gas and electricity spot prices λ_1^g and λ_1^e which corresponds to a correlation dependence of approximately 2 and 30 days probably due to the spikes components whereas the slower speed of mean reversion corresponds to a correlation dependence of 64 days and corresponds to the stochastic equilibrium or the normal variation of gas and electricity spot prices.

6.3 Theoretical background

In order to modelize heavy tails (and spikes) of stationary spot prices distribution, a natural idea is to consider an ergodic diffusion process like representation of deseasonalized spot prices.

In this section, we briefly recall how to construct a one dimensional process X solution of a stochastic differential equation with a prespecified continuous invariant density f. Throughout the sequel we assume that f is a strictly positive bounded continuous probability density on (l, r) (and zero outside (l, r)).

6.3.1 The general case

Let $\left(X_t\right)_{t\geq 0}$ the diffusion solution of the following stochastic differential equation (SDE)

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \quad X_0 \in (l, r), \tag{$E_{b,\sigma}$}$$

where $b : (l, r) \to \mathbb{R}$ and $\sigma : (l, r) \to \mathbb{R}$ are locally Lipschitz functions and that σ is not degenerate on (l, r) *i.e.* $\forall x \in (l, r), \sigma^2(x) > 0$. We introduce for the diffusion $\left(X_t\right)_{t\geq 0}$, the scale function $p : (l, r) \to \mathbb{R}$ defined for any $c \in (l, r)$ by

$$\forall x \in (l, r), \quad p(x) = \int_c^x \exp\left(-\int_c^y \frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\right)\mathrm{d}y,$$

and the speed measure density $m : (l, r) \to \mathbb{R}_+^*$ defined by

$$\forall x \in (l, r), \quad m(x) = \frac{2}{p'(x)\sigma^2(x)} = \frac{2}{\sigma^2(x)}\exp\left(\int_c^x \frac{2b(z)}{\sigma^2(z)}\mathrm{d}z\right). \tag{6.1}$$

We recall (see e.g. [48], [49]) that the process $\left(p(X_t^\zeta)\right)_{t\geq 0}$ with $\zeta = \inf\{t \geq 0, X_t = l \text{ or } X_t = r\}$ is a local martingale if and only if p is the scale function (unique up to an affine transformation). Moreover, if the diffusion $\left(X_t\right)_{t\geq 0}$ is positive recurrent, the stationary probability distribution ν defined on (l, r) satisfies

$$\nu(\mathrm{d}x) = Cm(x)\mathrm{d}x \quad \text{with} \quad C = \left(\int_l^r m(x)\mathrm{d}x\right)^{-1}.$$

This classical result is the key to construct a one-dimensional ergodic process that fits prescribed stationary probability distribution. For a more general result to construct an inhomogeneous Markov martingale process that has prespecified marginal density we refer to [63].

Proposition 6.3.1. *Let $b : (l, r) \to \mathbb{R}$ be a continuous drift function. Suppose that b and f satisfy the following conditions*

$$\forall x \in (l, r),\ \int_l^x b(y)f(y)\mathrm{d}y > 0, \quad \textit{and} \quad \int_l^r b(y)f(y)\mathrm{d}y = 0, \qquad (\mathcal{H}_B)$$

Then there exists a unique continuous diffusion function defined by

$$\forall x \in (l, r), \quad \sigma(x) = \sqrt{2\frac{\int_l^x b(y)f(y)\mathrm{d}y}{f(x)}},$$

such that $(E_{b,\sigma})$ has a unique solution $\left(X_t\right)_{t\geq 0}$, which is an ergodic diffusion process with stationary distribution ν satisfying $\nu(\mathrm{d}x) = f(x)\mathrm{d}x$.

Further details of the proof outlined below can be found in [15].

Proof. Let B be the function defined by $B(x) = \int_l^x b(y)f(y)\mathrm{d}y$. One checks easily that the scale function of $\left(X_t\right)_{t\geq 0}$ satisfies

$$\forall x \in (l, r), \quad p(x) = B(c)\int_c^x \frac{1}{B(y)}\mathrm{d}y.$$

One then obtains that $\lim_{x\to l} p(x) = -\infty$ and $\lim_{x\to r} p(x) = +\infty$.

On the other hand, the speed measure of $\left(X_t\right)_{t\geq 0}$ has density m that satisfies

$$\forall x \in (l, r), \quad m(x) = \frac{f(x)}{B(x)p'(x)} = \frac{f(x)}{B(c)}.$$

The normalized speed measure density is then equal to the probability density f.

To prove existence and uniqueness of the solution $\left(X_t\right)_{t\geq 0}$, one proves existence and uniqueness of the process $\left(p(X_t)\right)_{t\geq 0}$ satisfying a SDE without drift (see [48]). □

Corollary 6.3.2. *Let $b : x \in (l, r) \mapsto -\lambda(x-\mu)$ and assume that probability density f has expectation μ and finite variance. Then there exists a unique continuous diffusion function defined by*

$$\forall x \in (l, r), \quad \sigma(x) = \sqrt{\frac{\int_l^x 2\lambda(\mu - y)f(y)\mathrm{d}y}{f(x)}},$$

such that $(E_{b,\sigma})$ has a unique solution $\left(X_t\right)_{t\geq 0}$, which is an ergodic diffusion process with stationary distribution ν satisfying $\nu(\mathrm{d}x) = f(x)\mathrm{d}x$, and ACF given by

$$\forall t, \tau \geq 0, \quad \mathrm{Cor}(X_{t+\tau}, X_t) = e^{-\lambda\tau}.$$

The squared diffusion coefficients are explicitly known for a large number of commonly used probability diffusions. However, for some specific distributions, it is not possible to obtain a closed form of the diffusion coefficient. An approximation based on saddlepoint technique and the moment generating function (which is generaly known explicitly) is developed in [14].

6.3.2 Quasi-Saddlepoint approximation

We first recall that saddlepoint approximations are constructed by performing various operations on the moment generating function (MGF) of a random variable (see e.g. [18]). Let X be an absolutely continuous random variable with density f (with respect to the Lebesgue measure on (l,r)), moment generating function $M(t)$ and cumulant-generating function $\kappa(t) = \log M(t)$. Then the first-order saddlepoint density approximation to f is given by

$$\forall x \in (l,r), \quad \hat{f}(x) = \left(2\pi\kappa''(\hat{t}_x)\right)^{-1/2} e^{-\left(\hat{t}_x x - \kappa(\hat{t}_x)\right)},$$

where $t = \hat{t}_x$ is the (unique) solution to the saddlepoint equation $\kappa'(t) = x$, and primes denote derivatives. We assume that the probability density f has expectation μ, *i.e.* $\mu = \kappa'(0)$.

Considering the continuous differentiable function $\hat{t} : x \mapsto \hat{t}_x$, an integration by parts gives

$$\begin{aligned}\int_0^x \hat{t}(y)\mathrm{d}y &= \hat{t}(x)x - \int_0^x \hat{t}'(y)y\mathrm{d}y, \\ &= \hat{t}(x)x - \int_0^x \mathrm{d}\kappa(\hat{t}(y)),\end{aligned}$$

since $y = \kappa'(\hat{t}(y))$. The saddlepoint density $\hat{f}$ writes then

$$\forall x \in (l,r), \quad \hat{f}(x) = \left(2\pi\kappa''(\hat{t}(x))\right)^{-1/2} \exp\left(-\int_0^x \hat{t}(y)\mathrm{d}y\right). \tag{6.2}$$

To construct an ergodic process $\left(X_t\right)_{t\geq 0}$ solution of $(E_{b,\sigma})$ with prespicified stationary density $\hat{f}$, the exponential terms that appear in (6.2) and (6.1) suggest the relation $\frac{-2b}{\sigma^2} = t$. This construction is not exact but in [15] is proved that the speed density m of X is approximately propotional to the saddlepoint density $\hat{f}$. To be precise both $\sqrt{\kappa''(\hat{t}(x))}$ and $\sigma^2(x)$ are approximately proportional to $\kappa''(0)+\frac{1}{2}\kappa^{(3)}(0)\hat{t}(x)$ near the mean of the distribution. From now this normalized speed density m will be called the *quasi-saddlepoint* density approximation to f.

To summarize, if the saddlepoint function $\hat{t}$ is explicity known and efficiently computed, then we consider the diffusion with drift b, such that $b > 0$ on (l,μ) and $b < 0$ on (μ,r), and with diffusion coefficient

$$\forall x \in (l,r), \quad \sigma(x) = \sqrt{\frac{-2b(x)}{\hat{t}(x)}},$$

which is ergodic with stationary distribution $\tilde{f}(x) = \frac{c}{\sigma^2(x)} e^{-\left(x\hat{t}(x) - \kappa(\hat{t}(x))\right)}$ (where c is a normalizing factor), the quasi-saddlepoint density approximation to f (see [15] Theorem 3.1 for more details).

The following example will become useful later when we are going to modelize deseasonalized gas and electricity spot prices.

Example 5. The NIG-distribution *The normal-inverse Gaussian (NIG) distribution is a member of the class of generalized hyperbolic distributions (see e.g. [8]). The NIG density is given by*

$$f(x) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+(x-l)^2}\right)}{\pi\sqrt{\delta^2+(x-l)^2}} \times e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-l)}, \quad x \in \mathbb{R},$$

where $\beta \in \mathbb{R}$, $\alpha > |\beta|$, $\delta > 0$, $l \in \mathbb{R}$ *and* K_1 *is the the modified Bessel function of third order and index* 1. *Note that if* $X \sim \mathrm{NIG}(\alpha, \beta, \delta, l)$ *then its two first moments are*

$$\mathbb{E}[X] = l + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}} \quad \textit{and} \quad \mathrm{Var}(X) = \frac{\delta\alpha^2}{(\alpha^2-\beta^2)^{\frac{3}{2}}}.$$

The two parameters δ *and* l *determine respectivelly the scale and the location of the law, and the two parameters* α *and* β *determine the shape:* α *being responsible for the tail heavyness and* β *for the skewness (asymmetry).*

The cumulant-generating function κ *of the NIG distribution is defined for all* t *such that* $|\beta + t| < \alpha$ *by*

$$\kappa(t) = lt + \delta\left(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+t)^2}\right),$$

and the saddlepoint function is defined by

$$\forall x \in \mathbb{R}, \quad \hat{t}(x) = \frac{\alpha(x-l)}{\sqrt{\delta^2+(x-l)^2}} - \beta.$$

In order to have an Ornstein process solution of $(E_{b,\sigma})$ *with stationary density the quasi-saddlepoint density approximation* $\tilde{f}$ *to* f, *we consider the following drift and diffusion functions*

$$\forall x \in \mathbb{R}, \quad b(x) = -\lambda(x-\mu) \quad \textit{and} \quad \sigma^2(x) = \frac{2\lambda\sqrt{\delta^2+(x-l)^2}\,(x-\mu)}{\alpha(x-l) - \beta\sqrt{\delta^2+(x-l)^2}}, \tag{6.3}$$

with $\mu = l + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}}$.

6.4 Cross-commodity multi-factor model

In this section, we present two class of cross-commodity multi factor models: the geometric and the arithmetic class. Those two class are commonly used in stochastic modelling of commodity prices. The first one ensures the positivity of simulated spot prices. However, when dealing with forward contracts which have a delivery period or options pricing, the second one is analytically more tractable. Both class of models are based on stationary diffusion-type models analyzed in [15].

6.4.1 Proposed modelization

In order to represent the ACFs and CCF of gas and electricity deseasonalized spot prices, we are led to introduce stochastic processes that are sums of diffusions defined

by $(E_{b,\sigma})$. To be more precise, we focus on the following two factor modelization for the deseasonalized log spot prices Y^g and Y^e

$$Y_t^g = X_t^g + Z_t, \quad \text{and} \quad Y_t^e = X_t^e + Z_t, \tag{6.4}$$

where $\left(Z_t\right)_{t\geq 0}$, $\left(X_t^g\right)_{t\geq 0}$ and $\left(X_t^e\right)_{t\geq 0}$ are mutually independant processes defined as following:

- the process $\left(Z_t\right)_{t\geq 0}$ accounts for the stochastic equilibrium between both commodities with a slow rate of mean reversion $\lambda_z = \lambda_2^g = \lambda_2^e$. Thus, it represents the normal variation and will be defined by an Ornstein-Uhlenbeck process

$$\mathrm{d}Z_t = -\lambda_z Z_t \mathrm{d}t + \sigma_z \mathrm{d}W_t^z, \tag{6.5}$$

  with $\lambda_z > 0$ and $\sigma_z \in \mathbb{R}$. Note that Z is ergodic with the Gaussian invariant probability $\mathcal{N}\left(0, \sigma_z^2/2\lambda_z\right)$.

- the processes $\left(X_t^g\right)_{t\geq 0}$ and $\left(X_t^e\right)_{t\geq 0}$ represent the spikes component for each commodity. We modelize them by general Ornstein processes with high rate of mean reversion $\lambda_g = \lambda_1^g > 0$ and $\lambda_e = \lambda_2^e > 0$, namely

$$\mathrm{d}X_t^j = -\lambda_j \left(X_t^j - \mu_j\right) \mathrm{d}t + \sigma_j(X_t^j;\theta_j)\mathrm{d}W_t^j, \ j = g, e, \tag{6.6}$$

  where σ_j is a parametric diffusion function such that $\left(X_t^g\right)_{t\geq 0}$ is an ergodic diffusion with invariant probability $f^j(.,\theta_j)$.

Remark 8. *The following contruction can be extended to a more general multi-factor model. We can consider m general Ornstein processes and p Ornstein-Uhlenbeck processes so that*

$$Y^g(t) = \sum_{i=1}^{m} X_i^g(t) + \sum_{j=1}^{p} Z_j(t),$$
$$Y^e(t) = \sum_{i=1}^{m} X_i^e(t) + \sum_{j=1}^{p} Z_j(t),$$

where all processes are assumed to be mutually independent, i.e. driven by independent Wiener processes. We already observed that a two-factor model ($m = 1$ and $p = 1$) fits the ACFs and CCF well.

Proposition 6.4.1 (The correlation structures)**.** *Let Y^g, Y^e be the processes defined in* (6.4)*. Then, the ACFs of Y^g and Y^e with lag $\tau > 0$ are given by*

$$\rho^g(\tau) = \mathrm{Cor}\left(Y_{t+\tau}^g, Y_t^g\right) = \phi_g e^{-\lambda_g \tau} + (1-\phi_g)e^{-\lambda_z \tau},$$
$$\rho^e(\tau) = \mathrm{Cor}\left(Y_{t+\tau}^e, Y_t^e\right) = \phi_e e^{-\lambda_e \tau} + (1-\phi_e)e^{-\lambda_z \tau},$$

where

$$\phi_g = \frac{\mathrm{Var}\left(X^g(t)\right)}{\mathrm{Var}\left(Y^g(t)\right)}, \quad \textit{and} \quad \phi_e = \frac{\mathrm{Var}\left(X^e(t)\right)}{\mathrm{Var}\left(Y^e(t)\right)}.$$

The CCF with lag $\tau > 0$ is given by

$$\rho^{g,e}(\tau) := \mathrm{Cor}\left(Y_{t+\tau}^g, Y_t^e\right) = \phi_{g,e}\ e^{-\lambda_z \tau},$$

with, $\phi_{g,e} = \frac{\mathrm{Var}(Z(t))}{\sqrt{\mathrm{Var}(Y^g(t))\mathrm{Var}(Y^e(t))}}$.

From the definition of $\phi_{g,e}$, we find that $\sigma_z^2 = 2\lambda_z \phi_{g,e} \sqrt{\mathrm{Var}\left(Y^g(t)\right) \mathrm{Var}\left(Y^e(t)\right)}$, where the last term is the product of the two stationary variance of the two processes. Consequently, one can easily derive σ_z from the ACFs and CCF calibration.

6.4.2 Calibration

We propose a three-step calibration procedure for the model described above.

Step 1: Deseasonalizing spot prices

We fit the seasonality functions $g(t)$ and $e(t)$ defined in section 2.1 to the logarithmic spot prices. The parameters of the functions are estimated using the least squares approach. Now, we focus on the deseasonalized spot prices Y^g and Y^e defined by

$$Y^g(t) = \log\left(S^g(t)\right) - \log\left(g(t)\right) \quad \text{and} \quad Y^e(t) = \log\left(S^e(t)\right) - \log\left(e(t)\right).$$

One can consider the deseasonalized spot prices $e^{Y^g(t)}$ and $e^{Y^e(t)}$ instead of this geometric approach.

Step 2: ACFs and CCF

The least squares method consists in fitting the empirical ACFs $\rho^g(\tau)$, $\rho^e(\tau)$ and CCF $\rho^{g,e}(\tau)$ defined in section 2.4 to the empirical ones $\left(\tilde{\rho}^g(\tau)\right)_{\tau=1,\dots,l}$, $\left(\tilde{\rho}^e(\tau)\right)_{\tau=1,\dots,l}$, $\left(\tilde{\rho}^{g,e}(\tau)\right)_{\tau=1,\dots,l}$ in order to derive the three speeds of mean reversion λ_1^g, λ_1^e, λ_z with the diffusion coefficient σ_z of the stochastic equilibrium process Z. This can be done by minimizing the sum of squared differences, namely

$$\underset{\lambda_g,\lambda_e,\lambda_z,\sigma_z}{\arg\min} \sum_{\tau=1}^{l} \left(\left(\rho^g(\tau) - \tilde{\rho}^g(\tau)\right)^2 + \left(\rho^e(\tau) - \tilde{\rho}^e(\tau)\right)^2 + \left(\rho^{g,e}(\tau) - \tilde{\rho}^{g,e}(\tau)\right)^2 \right).$$

Stability tests showed that the estimates are robust with respect to small changes in the initial values of the parameters.

Step 3: Estimating the parameters of the spikes component

The final step consists in statiscally estimating the parameters θ_g of the invariant density $f^g(., \theta_g)$ of the process X^g and the parameters θ_e of the invariant density $f^e(., \theta_e)$ of the process X^e. For instance, if one decide to choose the quasi-saddlepoint approximation to the NIG density for f^g and f^e, there will be four parameters to fit for each density. The model proposed is a sum of diffusion processes and hence is not Markovian. Thus, the likelihood cannot be written down explicitly. To overcome this problem, we use the maximum likelihood estimation of order m ($m = 0$ or $m = 1$ in our case) method for stationary processes introduced in [5]. Strong consistency and a Central Limit Theorem are proved for such estimates. It consists in approximating the log-likelihood of the serie $(y_k^j)_{1 \le k \le n}$ $(j = g,\ e)$, where n is the number of sample points, by a sum whose generic term is the density function of Y_k^j conditional on the m most recent observations, for some $m \ge 0$, namely,

$$\ell_m^j(\theta) = \sum_{k=1}^{n} \log\left(h^j(y_k^j \mid y_k^{j,m}; \theta_j)\right), \tag{6.7}$$

where $y_k^{j,m} := (y_{k-m}^j, \cdots, y_{k-1}^j)$ and $h^j(. \mid y_k^{j,m}; \theta_j)$ is the conditional probability density function of Y_k^j given $Y_k^{j,m} = y_k^{j,m}$ for the parameters θ_j. Note that if $m = 0$, there is no conditioning and h^j is simply the marginal density of Y_k^j, $k = 1, \cdots, n$, which is the convolution of Z_k and X_k^j. We suppose that $(X_k^j)_{1\leq k\leq n}$ (resp. $(Z_k)_{1\leq k\leq n}$) is ergodic with stationary distribution $f^j(.;\theta_j)$ (resp. with Gaussian invariant probability $\mathcal{N}\left(0, \tilde{\sigma_Z}^2 := \sigma_z^2/2\lambda_z\right)$) so that the conditional probability density function is given by

$$h^j(y_k^j;\theta_j) = \int_{-\infty}^{+\infty} f^j\left(y_k^j - \frac{\sigma_Z}{\sqrt{\lambda_Z}}u;\theta_j\right)\frac{e^{-u^2}}{\sqrt{\pi}}\mathrm{d}u, \quad j = g, e. \tag{6.8}$$

Note that it corresponds to the case of $(Y_k^j)_{1\leq k\leq n}$ is independent and identically distributed random variables having the distribution of the stationary distribution of Y^j.

Numerically, the above integral can be approximated using a Gauss-Hermite quadrature method, namely

$$h^j(y_k^j;\theta_j) \approx \frac{1}{\sqrt{\pi}}\sum_{k=1}^{n} f^j\left(x - \frac{\sigma_Z}{\sqrt{\lambda_Z}}u_k;\theta_j\right)w_k, \quad j = g, e,$$

where $(u_k)_{1\leq k\leq n}$ are the roots of the Hermite polynomial P_n and $(w_k)_{1\leq k\leq n}$ are the associated weights given by

$$w_k = \frac{2^{n-1}n!\sqrt{\pi}}{n^2(P'_{n-1}(y_k))^2}, \quad k = 1, ..., n.$$

If $m = 1$, we need to compute the transition probability density $p_{Y_{k+1}^j|Y_k^j=y_k}(.;\theta_j)$ of $\left(Y_t^j\right)_{t\geq 0}$ for $j = g, e$. The two series $(X_k^j)_{1\leq k\leq n}$ and $(Z_k)_{1\leq k\leq n}$ are discrete observations of (6.6) and (6.5). Let g be a Borel bounded test function and $k \in 1, \cdots, n-1$, by conditioning we have

$$\begin{aligned}\mathbb{E}\left[g(Y_{k+1}^j) \mid Y_k^j = y_k\right] &= \int_{\mathbb{R}} \mathbb{E}\left[g(Y_{k+1}^j) \mid Y_k^j = y_k,\ Z_k = z\right]\mathbb{P}\left(Z_k = z \mid Y_k^j = y\right)\mathrm{d}z. \\ &= \int_{\mathbb{R}} \mathbb{E}\left[g(X_{k+1}^j + Z_{k+1}) \mid Y_k^j = y_k,\ Z_k = z\right]\mathbb{P}\left(Z_k = z \mid Y_k^j = y\right)\mathrm{d}z.\end{aligned}$$

Note that the two processes X^j and Z are independent so that if we denote by $p_{X_k^j}(x_k^j, .) := p_X(t_k, t_{k+1}, x_k^j, .)$ and $p_{Z_k}(z_k, .) := p_Z(t_k, t_{k+1}, z_k, .)$, the conditional probability density functions of X_{k+1}^j and Z_{k+1} given $Z_k = z,\ X_k^j = x_k^j$, the expectation $\mathbb{E}\left[g(X_{k+1}^j + Z_{k+1}) \mid Y_k = y_k^j,\ Z_k = z\right]$ is given by

$$\int_{\mathbb{R}} g(u)\int_{\mathbb{R}} p_{X_k^j}(y_k^j - z, v)p_{Z_k}(z, u - v)\mathrm{d}v\mathrm{d}u.$$

Moreover, we have

$$\mathbb{P}\left(Z_k = z \mid Y_k^j = y_k^j\right) = \frac{\mathbb{P}\left(Z_k = z,\ Y_k^j = y_k^j\right)}{\mathbb{P}\left(Y_k^j = y_k^j\right)} = \frac{\mathbb{P}\left(Z_k = z\right)\mathbb{P}\left(X_k^j = y_k^j - z\right)}{\mathbb{P}\left(Y_k^j = y_k^j\right)},$$

where,

$$\mathbb{P}\left(Y_k^j = y_k^j\right) = \int_{-\infty}^{+\infty} f^j\left(y_k^j - u; \theta_j\right) \frac{1}{\sqrt{2\pi}\tilde{\sigma}_Z} e^{-\frac{1}{2\tilde{\sigma}_Z^2}u^2} \mathrm{d}u,$$

and $\mathbb{P}\left(X_k^j = y_k^j - z\right) = f^j\left(y_k^j - z; \theta_j\right)$, $\mathbb{P}\left(Z_k = z\right) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_Z} e^{-\frac{1}{2\tilde{\sigma}_Z^2}z^2}$. Finally, one easily identifies the transition probability density $p_{Y_{k+1}^j | Y_k^j = y_k^j}(y; \theta_j)$ which is given by

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} p_{X_k^j}(y_k^j - z, v) p_{Z_k}(z, u - v) \mathrm{d}v \right) \frac{1}{\sqrt{2\pi}\tilde{\sigma}_Z} \frac{f^j\left(y_k^j - z; \theta_j\right) e^{-\frac{1}{2\tilde{\sigma}_Z^2}z^2}}{\mathbb{P}\left(Y_k^j = y_k^j\right)} \mathrm{d}u.$$

Note that we have $p_{Z_k}(z_k, z) = \frac{1}{\sqrt{2\pi}\bar{\sigma}_Z} e^{-\frac{1}{2\bar{\sigma}_Z^2}z^2}$, with $\bar{\sigma}_Z = \sigma_Z \sqrt{\frac{1 - e^{-2\lambda_Z \Delta}}{2\lambda_Z}}$ using an exact scheme of the Ornstein-Uhlenbeck process $(Z_k)_{1 \leq k \leq n}$ of step $\Delta > 0$, namely

$$Z_{k+1} = e^{-\lambda_z \Delta} Z_k + \sigma_Z \sqrt{\frac{1 - e^{-2\lambda_Z \Delta}}{2\lambda_Z}} G_{k+1}^z, \tag{6.9}$$

where $(G_k^z)_{k \geq 1}$ is a sequence of i.i.d. standard normal random variables. However, in most cases, there is no closed expression for $p_{X_k^j}(x_k^j, .)$. To overcome this problem one solution is to consider the transition probability density function $p_{\bar{X}_k^j}(x_k^j, .)$ of the Euler scheme $\left(\bar{X}_k^j\right)_{k \geq 0}$

$$\bar{X}_{k+1}^j = e^{-\lambda_j \Delta} \bar{X}_k^j + \mu_j\left(1 - e^{-\lambda_j \Delta}\right) + \sigma_j\left(\bar{X}_k^j; \theta_j\right) \sqrt{\frac{1 - e^{-2\lambda_j \Delta}}{2\lambda_j}} G_{k+1}^j, \ k \geq 0 \tag{6.10}$$

where $\left(G_k^j\right)_{k \geq 1}$ is a sequence of i.i.d. standard normal random variables independent of $(G_k^z)_{k \geq 1}$. Consequently, $p_{\bar{X}_k^j}(x_k^j, .) = \frac{1}{\sqrt{2\pi}\bar{\sigma}_j(x_k^j; \theta_j)} e^{-\frac{1}{2\bar{\sigma}_j^2(x_k^j; \theta_j)}x^2}$, with $\bar{\sigma}_j(x_k^j; \theta_j) = \sigma_j\left(\bar{x}_k^j; \theta_j\right) \sqrt{\frac{1 - e^{-2\lambda_j \Delta}}{2\lambda_j}}$ so that we have

$$\int_{\mathbb{R}} p_{\bar{X}_k^j}(y_k^j - z, v) p_{Z_k}(z, u - v) \mathrm{d}v = \frac{1}{\sqrt{2\pi}\tilde{\sigma}(y_k^j, z; \theta^j)} e^{-\frac{1}{2\tilde{\sigma}^2(y_k^j, z; \theta^j)}\left(u - m_k^j\right)^2},$$

where for $k \in \{1, \cdots, n\}$, $m_k^j = e^{-\lambda_Z \Delta} z + e^{-\lambda_j \Delta}(y_k^j - z) + \mu_j(1 - e^{-\lambda_j \Delta})$ and $\tilde{\sigma}^2(y_k^j, z; \theta^j) = \bar{\sigma}_j^2\left(y_k^j - z; \theta_j\right) + \bar{\sigma}_Z^2$.

Remark 9. *In [28], a transition probability density function based on Milstein scheme is used. In [52], a gaussian transition probability density function with Taylor expansions is used to propose an efficient estimator for θ_j.*

The method of maximum likelihood of order m estimates $\hat{\theta}_{j,m}$ by finding the value of θ_j that maximizes (6.7) using standard numerical optimization procedure.

6.5 Simulation and application

6.5.1 Empirical results on Powernext and NBP spot prices

In this section, we perform the calibration procedure on electricity spot prices coming from the Powernext market and on gas spot prices at the NBP. Then, we perform a simulation with estimated parameters over the same period. To avoid negative prices, we choose to represent spot prices by an arithmetic model, namely

$$S^g(t) = g(t) \times e^{X^g(t)+Z(t)}, \tag{6.11}$$

$$S^e(t) = e(t) \times e^{X^e(t)+Z(t)}, \tag{6.12}$$

where $g(t)$, $e(t)$ are the trend and seasonality functions defined in Section 6.2.1, X^g, X^e are solutions of $(E_{b,\sigma})$ with b and σ defined in (6.3) and Z is a Gaussian Ornstein-Uhlenbeck process solution of (6.5).

We choose the NIG distribution for those two processes in order to capture the heavy tails behavior observed on data, *i.e.* large values with low probability that cannot be obtained by a Gaussian process. We observed that the quasi-saddlepoint approximation of the NIG-distribution is well suited to represent the two spike components. One can choose another distribution and devise the same calibration process as in the previous section. The results of steps 1 and 2 of the calibration procedure are reported in Figure 6.1 and the quality of the ACFs and CCF fits is represented in Figure 6.5. Now, we proceed to the estimation of the four parameters $\theta_g = (\alpha_g, \beta_g, \delta_g, l_g)$ of the process X^g and the four parameters $\theta_e = (\alpha_e, \beta_e, \delta_e, l_e)$ of the process X^e using the maximum likelihood estimation method described in the previous section on the deseasonalized spot prices. We observed that the maximum likelihood estimation method of order 0 is more robust and gives better results than the one of order 1[4]. The initial parameters are set to $(1, 0, 1, 0)$ for both components.

The algorithms converged quickly. The diffusion coefficient functions $\tilde{\sigma}_j(., \theta_j)$, $j = g, e$, with the fitted parameters, are documented in Figure 6.6. We see that the shape of the diffuion coefficients are quite similar for the gas and electricity spot deseasonalized spot prices. Spikes are obtained when the processes Y^g and Y^e are far from their mean by clusters of volatility, *i.e.* periods of high volatility. As we see, large values are more likely and the asymmetry is more pronounced for electricity spot prices than for gas spot prices. We clearly see spikes as cluster of volatility are more probable and more intense for electricity deseasonalized spot prices than for gas deseasonalized spot prices.

In order to simulate price trajectories, we consider the Euler-Maruyama schemes defined by (6.10) and (6.9). If one is concerned by estimating some quantities (for instance quantiles) on only one trajectory then one should replace the above Euler schemes of X^g and X^e with their respective Milstein schemes $\tilde{X}^g$ and $\tilde{X}^e$ in order to achieve a smaller strong error rate. It consists in devising the following schemes

[4]Fitted parameters of order 1 are: $\alpha_g = 0.76$, $\beta_g = 7.8e-2$, $\delta_g = 7.8e-4$, $l_g = -0.11$ and $\alpha_e = 1.56$, $\beta_e = 0.34$, $\delta_e = 1.1e-2$, $l_e = 0.16$

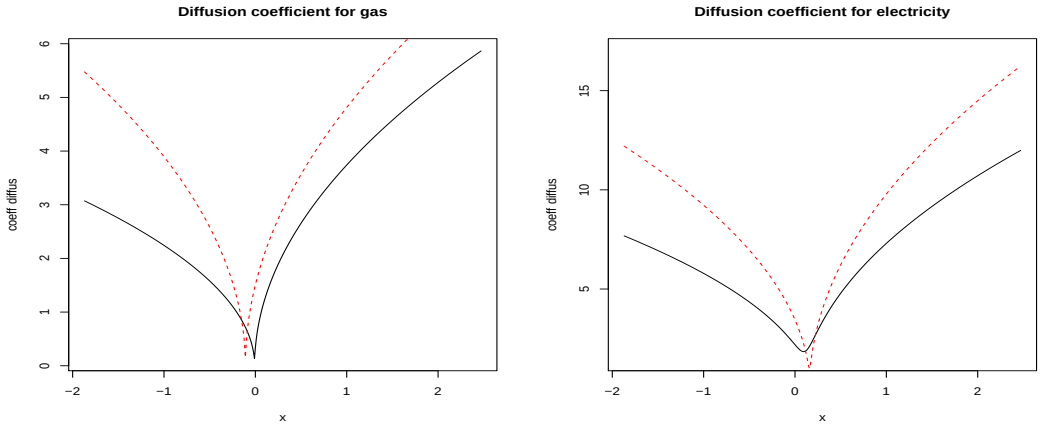


Figure 6.6: Squared diffusion coefficients using fitted parameters with maximum likelihood estimation of order 0 (normal lines) and of order 1 (dashed lines). Fitted parameters of order 0 are: $\alpha_g = 1.93$, $\beta_g = 0.90$, $\delta_g = 2.25e-3$, $l_g = -8.8e-3$ and $\alpha_e = 3.49$, $\beta_e = 1.24$, $\delta_e = 0.08$, $l_e = 0.11$.

for $j = g, e$,

$$\tilde{X}^j_{t_{k+1}} = e^{-\lambda_j(t_{k+1}-t_k)}\left(\tilde{X}^j_{t_k} + \left(\mu_j\lambda_j - \frac{1}{2}\sigma_j\sigma_j{}'(\tilde{X}^j_{t_k};\theta_j)\right)\Delta\right) + \sigma_j\left(\tilde{X}^j_{t_k};\theta_j\right)\sqrt{\frac{1-e^{-2\lambda_j\Delta}}{2\lambda_j}}G^j_{k+1} + \frac{1}{2}\sigma_j\sigma_j{}'(\tilde{X}^j_{t_k};\theta_j)\left(G^j_{k+1}\right)^2,\ \tilde{X}^j_0 = x^j_0,$$

where $\sigma_j{}'$ is the first derivative of σ_j.

In the following simulations, we consider Milstein schemes of step $t_k = k\Delta$, with $\Delta = \frac{1}{252}$. Next, we add to the simulated processes the two seasonality functions. In Figure 6.7, the simulated deseasonalized spot prices are represented. We see that both commodities are strongly linked and that the model mimics the statistical behaviour of the deseasonalized spot prices. In Figure 6.8, the simulated spot prices are represented. In Figure 6.9, both simulated and historical ACFs and CCF are plotted. We clearly see that the model reproduces the correlation structures.

6.5.2 Application: measuring risk of a cross-commodity portfolio

In this section, we aim at measuring the risk of a portfolio composed of a short position in a power plant that produces electricity from gas day by day $t_1 < \ldots < t_N$ for several maturities $T = t_N =$ 6 months, 1 year and 3 years. The loss at time 0 of the portfolio with a time horizon T can be written

$$L_T = \sum_{k=1}^{N} e^{-rt_k}\left(S^e_{t_k} - h_R S^g_{t_k} - C\right)_+ - P^c_T,$$

where $r = 5\%$ is the annual interest rate, $h_R = 3$ denotes the Heat Rate, $C = 5$ €/MWh denotes the generation costs and where P^c_T is an estimation of the price of

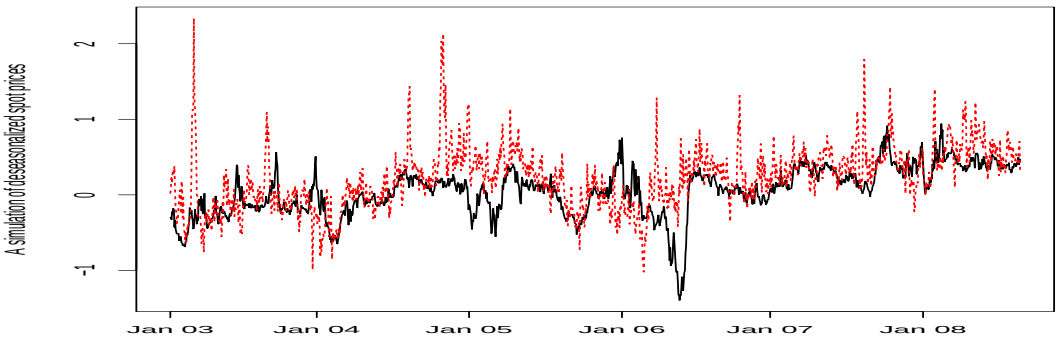


Figure 6.7: A simulation of gas (normal line) and electricity (dotted line) deseasonalized spot prices.

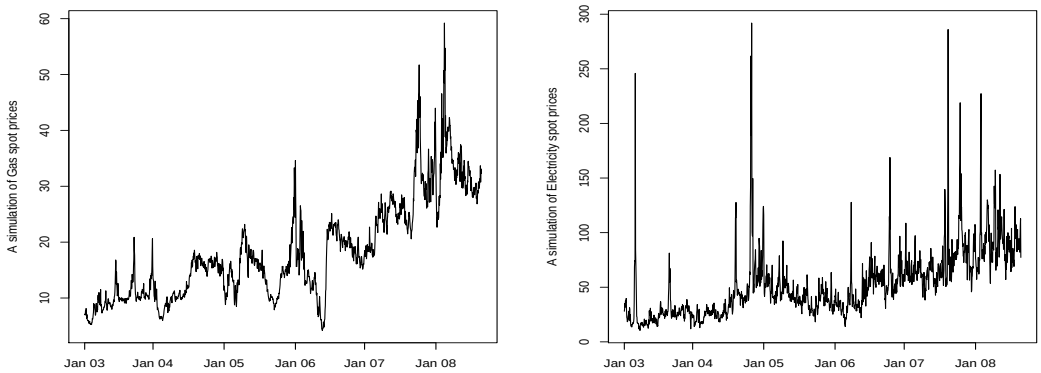


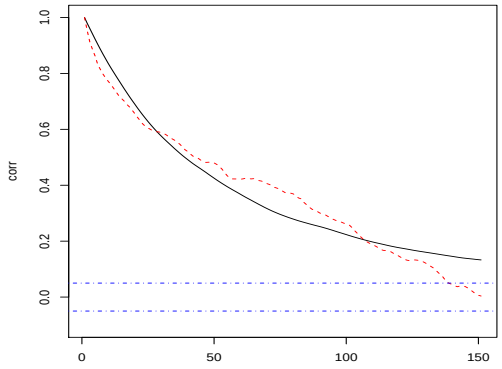
Figure 6.8: Simulated Electricity spot prices on the Powernext market on the left and Gas spot prices at the NBP on the right for the period 14 January 2003 till 20 August 2008.

the option on the power plant obtained by a crude Monte Carlo simulation, namely

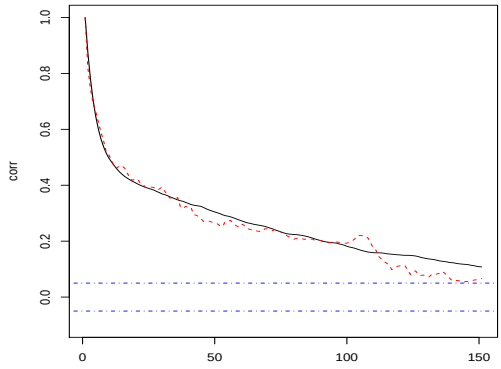
$$P_T^c \approx \sum_{k=1}^{N} e^{-rt_k} \mathbb{E}\left[\left(S_{t_k}^e - h_R S_{t_k}^g - C\right)_+\right].$$

Since gas and electricity markets are incomplete, we price and estimate risk measures under the historical probability. In order to measure the risk, we consider the Value-at-Risk (VaR), which is certainly the most commonly used risk measures in the context of risk management. By definition, the Value-at-Risk at level $\alpha \in (0,1)$ (VaR_α) of a given portfolio is the lowest amount not exceeded by its loss with probability α. In this example, we set $\alpha = 95\%$. Actually, for the considered portfolio, the VaR_α is the unique solution ξ of the equation

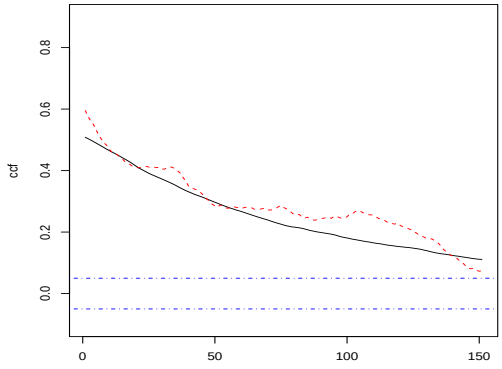
$$\mathbb{P}\left(L_T \leq \xi\right) = \alpha.$$



(a) ACF of one simulation of Y^g



(b) ACF of one simulation of Y^e



(c) CCF of the simulations

Figure 6.9: ACFs and CCF of simulated gas and electricity spot prices (normal lines) with the historical ACFs and CCF (dotted lines).

The portfolio's VaR_α is just a quantile of its loss and is interpreted as a reasonable worst case level.

Now, we are interested in measuring the impact of the proposed model for gas and electricity spot prices on the portfolio's VaR. In order to do that, we consider three different models:

- Case 1: the mean-reverting cross-commodity model (in its geometric form) proposed in this paper and defined by (6.11) and (6.12). It modelizes typical features of gas and electricity spot prices likes spikes and the long term dependency.

- Case 2: a slight modification of the previous model in which we do not take into account the dependence of the two energy spot prices. To be more precise,

we consider the following model specification

$$S^g(t) = g(t) \times e^{X^g(t)+Z^g(t)},$$
$$S^e(t) = e(t) \times e^{X^e(t)+Z^e(t)},$$

where X^g and X^e are solutions of $(E_{b,\sigma})$ with b and σ defined in (6.3), and where Z^g, Z^e are two independant Gaussian OU processes solution of (6.5). By this model, we want to measure the impact on the VaR_α of the long term dependency modeling. The calibration process is slightly modified since S^g and S^e are now independent. The step 2 is replaced by two different minimizations corresponding to each ACF. Steps 1 and 3 remain unchanged.

- Case 3: a slight modification of the case 1 in which we do not modelize the spikes feature. To be more precise, we replace the NIG-distributed processes by Gaussian Ornstein-Uhlenbeck processes, namely

$$S^g(t) = g(t) \times e^{Z^g(t)+Z(t)},$$
$$S^e(t) = e(t) \times e^{Z^e(t)+Z(t)},$$

where Z^g, Z^e, Z are three different Gaussian OU processes solution of (6.5). By this model, we want to quantify the impact on the VaR_α of the spike feature of gas and electricity spot prices.

In each case, we estimate P_T^c and the VaR_α using 10 000 Monte Carlo simulations. We devise Euler schemes of step $t_k = k\Delta$ with $\Delta = \frac{1}{252}$. In order to estimate the VaR_α, we use the inversion of the simulated empirical distribution function.

Remark 10. *Since gas and electricity spot prices are sums of diffusion processes solution of $(E_{b,\sigma})$, one can easily use the method investigated in Chapter 3 to estimate the VaR_α and other risk measures. It is based on stochastic approximation algorithms with an adaptive variance reduction tool (unconstrained importance sampling algorithm). The method is known to achieve good variance reduction when $\alpha \approx 1$ as it is often the case. For the sake of simplicity, we only considered the classical method based on the inversion of the empirical distribution function.*

The results are summarized in Tables 6.1. Note that for each case, the estimations are computed using the same pseudo-random number generator initialized with the same *seed*. The number in parentheses refers to the 95% confidence level.

We observe that there are slight differences in terms of the price P_T^c between the case 1 and 2 but huge differences in terms of risk. Taking into account the long term correlation between gas and electricity spot prices can reduce substantially the risk of this portfolio. Modeling independently each energy spot prices leads to an overestimation of the VaR_α of the portfolio's loss. The results obtained by using the model investigated in case 3 shows that introducing the spikes behavior into the model can increase greatly both P_T^c and the risk of the portfolio. We also estimated the same quantities using the arithmetic version of the three models presented above. We obviously obtained different values from the ones presented but the same conclusions hold: modeling adequatly the cross correlation between gas and electricity spot prices reduces the risk of portfolio whereas modeling adequatly the spiky behavior of both commodities increases greatly the price of the option and the risk associated to the portfolio.

	Maturity	P_T^c	($\pm$Error)	VaR_α
Case 1 (Proposed model)	6 months	83.3	($\pm$3.3)	262.4
	1 year	220.1	($\pm$5.5)	495.4
	3 years	745.0	($\pm$11.2)	1081.0
Case 2 (No cross-correlation)	6 months	51.2	($\pm$2.9)	250.1
	1 year	222.6	($\pm$8.4)	880.2
	3 years	850.6	($\pm$21.3)	2213.1
Case 3 (Gaussian model)	6 months	32.9	($\pm$1.1)	107.7
	1 year	129.8	($\pm$2.7)	275.9
	3 years	437.1	($\pm$5.8)	565.5

Table 6.1: Estimation of the price of the Power plant and the VaR_α of the portfolio.

Bibliography

[1] B. Arouna : Adaptive monte carlo method, a variance reduction technique. *Monte Carlo Methods and Applications*, 10(1):1–24, 2004.

[2] B. Arouna et O. Bardou : Efficient variance reduction for functionals of diffusions by relative entropy. Rapport technique, CERMICS-ENPC, 2004.

[3] P. Artzner, F. Delbaen, J. M. Eber et D. Heath : Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[4] M. Avellaneda : Minimum-relative-entropy calibration of asset pricing models. *International Journal of Theoretical and Applied Finance*, 1:447–472, 1998.

[5] A. Azzalini : Maximum likelihood estimation of order m for stationary stochastic processes. *Biometrika*, 70:381–387, 1983.

[6] O. Bardou, N. Frikha et G. Pagès : Computing var and cvar using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009.

[7] M. T. Barlow : A diffusion model for electricity prices. *Math. Finance*, 12(4):287–298, 2002.

[8] Ole E. Barndorff-Nielsen : Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.*, 24(1):1–13, 1997.

[9] P. Barrieu et N. El Karoui : Optimal derivatives design under dynamic risk measures. *Mathematics of Finance*, A.M.S. Proceedings 351, 2004.

[10] F.E. Benth, J. Kallsen et T. Meyer-Brandis : A non-Gaussian Ornstein-Uhlenbeck process for electricity spot price modeling and derivatives pricing. *Appl. Math. Finance*, 14(2):153–169, 2007.

[11] F.E. Benth et P.C. Kettler : Dynamic copula models for the spark spread. Rapport technique, E-print no. 14, Department of Mathematics, University of Oslo, 2006.

[12] F.E. Benth et R. Kufakunesu : Pricing of exotic energy derivatives based on arithmetic spot models. Rapport technique, 2007.

[13] M. Benveniste, M. Metivier et P. Priouret : *Adaptive algorithms and stochastic approximation*. Springer-Verlag, 1990.

[14] B.M. Bibby, I.M. Skovgaard et M. Sørensen : Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, 11(2):191–220, 2005.

[15] B.M. Bibby et M. Sørensen : A hyperbolic diffusion model for stock prices. *Finance Stoch.*, 1(1):25–41, 1997.

[16] V.S. Borkar : Stochastic approximation with two time scales. *Systems Control Lett.*, 29:291–294, 1997.

[17] M. Britten-Jones et S.M. Schaefer : Non linear value-at-risk. *European Finance Review*, 2:161–187, 1999.

[18] Ronald W. Butler : *Saddlepoint approximations with applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2007.

[19] L. De Clerck : A proof of niederreiter's conjecture concerning error bounds for quasi-monte carlo integration. *Adv. in Appl. Math.*, 2:1–6, 1981.

[20] S.J. Deng et W. Jiang : Levy process-driven mean-reverting electricity price model: the marginal distribution analysis. *Decision Support Systems*, 40(3-4):483–494, 2005.

[21] D. Duffie et J. Pan : Analytical value-at-risk with jumps and credit risk. *Finance and Stochastics*, 5(2):155–180, 2001.

[22] M. Duflo : *Iterative random models*. Springer-Verlag, 1988.

[23] M. Duflo : *Algorithmes Stochastiques*. Springer, Berlin, 1996.

[24] D. Dufresne et F.J. Vázquez-Abad : Accelerated simulation for pricing asian options. *Proceedings of the 1998 Winter Simulation Conference*, pages 1493–1500, 1998.

[25] D. Egloff et M. Leippold : Quantile estimation with adaptive importance sampling. *Electronic copy available at: http://ssrn.com/abstract=1002631*, 2007.

[26] N. El Karoui et M.-C. Quenez : Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM J. Control and Optim.*, 33:29–66, 1995.

[27] N. El Karoui et R. Rouge : Pricing via utility maximsation and entropy. *Mathematical Finance*, 10:250–276, 2000.

[28] O. Elerian : A note on the existence of a closed form conditional transition density for milstein scheme. *Working Paper, Nuffield College, Oxford University*, 1998.

[29] V. Evstigneev, I. : Measurable selection and dynamic programming. *MAthematics of operations research*, 1:267–275, 1976.

[30] H. Foellmer et P. Leukert : Quantile hedging. *Finance and Stochastics*, 3:251–273, 1999.

[31] H. Foellmer et A. Schied : Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[32] H. Foellmer et M. Schweizer : Hedging of contingent claims under incomplete information. *Applied Stochastic Analysis*, pages 389–414, 1991.

[33] H. Foellmer et D. Sondermann : Hedging of non-redundant contingent claims. *Contributions to Mathematical Economics*, pages 205–223, 1986.

[34] M. Fritteli : The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10:39–52, 2002.

[35] M.C. Fu et Y. Su : Optimal importance sampling in securities pricing. *Journal of Computational Finance*, 5(4):27–50, 2000.

[36] H. Geman : *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy*. Wiley & Sons, 2005.

[37] H. Geman et A. Roncoroni : Understanding the fine structure of electricity prices. *The Journal of Business*, 79(3):1225–1261, 2006.

[38] P. Glasserman, P. Heidelberger et P. Shahabuddin : Importance sampling and stratification for value-at-risk. *Computational Finance*, pages 7–24, 1999.

[39] P. Glasserman, P. Heidelberger et P. Shahabuddin : Variance reduction techniques for estimating value-at-risk. *Management Science*, 47:1349–1364, 1999.

[40] P. Glasserman, P. Heidelberger et P. Shahabuddin : Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12:239–270, 2002.

[41] P. Glasserman et Y. Wang : Counterexamples in importance sampling for large deviation probabilities. *Annals of Applied Probability*, 7(3):731–746, 1997.

[42] E. Gobet : Weak approximation of killed diffusion using euler schemes. *Stochastic Processes and their Applications*, 87(2):167–197, 2000.

[43] D. Harrison, J. et Kreps : Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, pages 381–408, 1979.

[44] S. D. Hodges et A. Neuberger : Optimal replication of contingent claims under transaction costs. *The Review of Futures Markets*, 8:222–239, 1989.

[45] A.B. Juditsky et B.T Polyak : Acceleration of stochastic approximation by averaging. *Journal on Control and optimization*, 30(4):838–855, 1992.

[46] T. Kanamura et K. Ōhashi : A structural model for electricity prices with spikes: Measurement of spike risk and optimal policies for hydropower plant operation. *Energy Economics*, 29(5):1010–1032, 2007.

[47] I. Karatzas, J.P. Lehoczky, S. Shreve et G. Xu : Martingale and duality methods for utility maximization in an incomplete market. *SIAM J. Control and Optimization*, 29:702–730, 1991.

[48] I. Karatzas et S.E. Shreve : *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics, 113. New York etc.: Springer-Verlag. xxiii, 470 p. DM 68.00/pbk , 1991.

[49] S. Karlin et H.M. Taylor : *A second course in stochastic processes.* New York etc.: Academic Press, A Subsidiary of Harcourt Brace Jovanovich, Publishers. XVIII, 542 p. $ 35.00 , 1981.

[50] R. Kawai : Optimal importance sampling parameter search for lévy processes via stochastic approximation. *SIAM Journal on Numerical Analysis*, 47(1):293–307, 2008.

[51] A. Kemna et T. Vorst : A pricing method for options based on average asset values. *Journal of Banking and Finance*, 14:113–130, 1990.

[52] M. Kessler : Estimation of an ergodic diffusion from discrete observations. *Journal of Statistics*, 24:211–229, 1997.

[53] J. Kiefer et J. Wolfowitz : Stochastic estimation of the maximum of a regression function. *Ann. Math. Statis.*, 23:462–466, 1952.

[54] V.R. Konda et J.N. Tsitsiklis : Convergence rate of linear two-time-scale stochastic approximation. *Annals Applied Probability*, 14(2):796–819, 2004.

[55] D. P. Kroese et R. Y. Rubinstein : *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte Carlo Simulation and Machine Learning.* Springer, New York, 2004.

[56] L. Kuipers et H. Niederreiter : *Uniform Distribution of Sequences.* John Wiley, 1974.

[57] H. J. Kushner et G. Yin : *Stochastic Approximation and Recursive algorithms and Applications.* Springer, 2003.

[58] B. Lapeyre, G. Pagès et K. Sab : Sequences with low discrepancy. generalization and application to robbins-monro algorithm. *Statistics*, 21(2):251–272, 1990.

[59] B. Lapeyre et E. Temam : Competitive monte carlo methods for the pricing of asian options. *Journal of Computational Finance*, 5:39–57, 2003.

[60] S. Laruelle et G. Pagès : Stochastic approximation with averaging innovation. *Preprint*, 2010.

[61] J. Lelong : *Algorithmes stochastiques et Options parisiennes.* Thèse de doctorat, ENPC, 2007.

[62] V. Lemaire et G. Pagès : Unconstrained recursive importance sampling. *To appear in Annals of Applied Probability*, 2008.

[63] D.B. Madan et M. Yor : Making Markov martingales meet marginals: with explicit constructions. *Bernoulli*, 8(4):509–536, 2002.

[64] W. Margrabe : The value of an option to exchange one asset for another. *Journal of Finance*, 33(1):177–186, 1978.

[65] R.C. Merton : Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3:373–413, 1971.

[66] T. Meyer-Brandis et P. Tankov : Multi-factor jump-diffusion models of electricity prices. *Int. J. Theor. Appl. Finance*, 11(5):503–528, 2008.

[67] Y. Miyahara : Pricing model and related estimation problems. *Asia-Pacific Financial Markets*, 8:45–60, 2001.

[68] A. Mokkadem et M. Pelletier : Convergence rate and averaging of non linear two-time-scale stochastic approximation algorithms. *The Annals of Applied Probability*, 16(3):1671–1702, 2006.

[69] M.B. Nevelson et R.Z. Khasinskii : Stochastic approximation and recursive estimation. *Translation of Math. Monographs American Mathematical Society*, 47, 1976.

[70] H. Niederreiter : *Random number generation and quasi-Monte Carlo methods, volume 63 of CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), 1992.

[71] H. Niederreiter et J. M. Wills : Diskrepanz und distanz von maβen bezüglich konvexer und jordanscher mengen. *Math. Z.*, 144:125–134, 1975.

[72] G. Pagès : A space vector quantization method for numerical integration. *J. Computational and Applied Mathematics*, 89:1–38, 1998.

[73] M. Pelletier : Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39:49–72, 2000.

[74] Mariane Pelletier : Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39(1):49–72 (electronic), 2000.

[75] D. Proïnov, P. : Discrepancy and integration of continuous functions. *Journal of Approximation Theory*, 52(2):1608–1622, 1988.

[76] H. Robbins et S. Monro : A stochastic approximation method. *Ann. Math. Statis.*, 22:400–407, 1951.

[77] R. T. Rockafellar et S. Uryasev : Optimization of conditional value-at-risk. *Journal of Risk*, 2:493–517, 2000.

[78] R.T. Rockafellar et S. Uryasev : Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

[79] L.C.G. Rogers et D. Williams : *Diffusions, Markov Processes and Martingales*. Cambridge Mathematical Library, 1986.

[80] C. Rouvinez : Going greek with var. *Risk*, 10(2):57–65, 1997.

[81] D. Ruppert : Stochastic approximation. *Handbook of Sequential Analysis*, pages 503–529, 1991.

[82] E.S. Schwartz : The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of finance*, pages 923–973, 1997.

[83] M. Schweizer : Option hedging for semimartingales. *Stochastic Processes and Their Applications*, pages 339–363, 1991.

[84] R. J. Serfling : *Approximation Theorems for Mathematical Statistics*. Wiley, New York, 1980.

[85] P. Villaplana : Pricing power derivatives: A two-factor jump-diffusion approach. Conference Paper, European Finance Association. *In 30th Annual Meeting*, 2003.

Résumé

Cette thèse est consacrée à des problématiques numériques probabilistes liées à la modélisation, au contrôle et à la gestion du risque et motivées par des applications dans les marchés de l'énergie. Le principal outil utilisé est la théorie des algorithmes stochastiques et des méthodes de simulation. Cette thèse se compose de trois parties. La première est dévouée à l'estimation de deux mesures de risque de la distribution L des pertes d'un portefeuille: la Value-at-Risk (VaR) et la Conditional Value-at-Risk (CVaR). Cette estimation est effectuée à l'aide d'un algorithme stochastique combiné avec une méthode de réduction de variance adaptative. La première partie de ce chapitre traite du cas de la dimension finie, la deuxième étend la première au cas d'une fonction de la trajectoire d'un processus et la dernière traite du cas des suites à discrépance faible. Le deuxième chapitre est dédié à des méthodes de couverture du risque en CVaR dans un marché incomplet opérant à temps discret à l'aide d'algorithmes stochastiques et de quantification vectorielle optimale. Des résultats théoriques sur la couverture en CVaR sont présentés puis les aspects numériques sont abordés dans un cadre markovien. La dernière partie est consacrée à la modélisation conjointe des prix des contrats spot Gaz et l'Électricité. Le modèle multi-facteur présenté repose sur des processus d'Ornstein stationnaires à coefficient de diffusion paramétrique.

Mot-clefs

Approximation stochastique, Marchés de l'énergie, Echantillonnage préférentiel, Value-at-Risk, Conditional Value-at-Risk, Couverture du risque, Modèle multi-facteur, Processus stationnaire

Title

Contribution to modeling and dynamic risk hedging of energy markets

Abstract

This thesis is concerned with probabilistic numerical problems about modeling, risk control and risk hedging motivated by applications to energy markets. The main tool is based on stochastic approximation and simulation methods. This thesis consists of three parts. The first one is devoted to the computation of two risk measures of the portfolio loss distribution L: the Value-at-Risk (VaR) and the Conditional Value-at-Risk (CVaR). This computation uses a stochastic algorithm combined with an adaptive variance reduction technique. The first part of this chapter deals with the finite dimensional case, the second part extends the results of the first part to the case of a path-dependency process and the last one deals low discrepancy sequences. The second chapter is devoted with risk minimizing hedging strategies in an incomplete market operating in discrete time using quantization based stochastic approximation. Theoretical results on CVaR hedging are presented then numerical aspects are adressed in a Markovian framework. The last part deals with joint modeling of Gas and Electricity spot prices. The multi-factor model presented is based on stationnary Ornstein process with parameterized diffusion coefficient.

Keywords

Stochastic approximation, Energy markets, Importance Sampling, Value-at-Risk, Conditional Value-at-Risk, Risk hedging, Multi-factor model, stationnary process.

Laboratoire de Probabilités et Modèles Aléatoires,
CNRS-UMR 7599, UFR de Mathématiques, case 7012

Université Paris 6, Pierre & Marie Curie
4, place Jussieu, 75251 Paris Cedex 05.